# MOLECULAR AND GENETIC ANALYSIS OF *Dominant hemimelia* AND *Hemimelic extra toes*; MOUSE MUTATIONS WHICH AFFECT LIMB DEVELOPMENT

**Jeremy P. Allen**

**PhD**
**The University of Edinburgh**
**1997**

# DECLARATION

I declare that:

a) this thesis has been composed by myself

b) that the work is my own, except where otherwise stated

Jeremy P. Allen

September 1997

# ABSTRACT

The mouse developmental mutation *Dominant hemimelia, Dh*, is distinguished by pleiotrophic effects which are confined to the posterior half of the animal. *Dh* causes skeletal abnormalities of the hindlimbs characterised by tibial hemimelia, luxation, poly/oligodactyly and fusion or loss of pelvic girdle bone elements. The number of pre-sacral vertebrae, sternebrae and ribs are also reduced and visceral defects include asplenia. To understand the molecular basis of the *Dh* phenotype this thesis, in part, describes attempts to identify the *Dh* gene through the technology of positional cloning. Previous analysis of an intraspecific backcross has localised *Dh* to a region of 1.2cM on chromosome 1. To characterise this region a genomic map and yeast artificial chromosome (YAC) contig covering 2.4Mb has been constructed. The location of two genes which are genetically inseparable from *Dh* are described; G*li2*, a member of the *Gli* family of zinc finger genes and *Inhibinβb*, a member of the *TGFβ* superfamily. Both were analysed for mutation in *Dh* animals but none were detected.

A relationship between *Dh* and a second mutation *Hemimelic extra toes*, *Hx*, which presents a very similar limb phenotype has been proposed by virtue of the two genes residing in apparent paralogous linkage groups on chromosomes 1 and 5 respectively. This suggests a common ancestry for the two genes. The candidacy of *Gli2* for *Dh* raised the possibility that a novel *Gli* gene was involved in *Hx*. Identification of such a novel *Gli* gene was attempted by degenerate PCR. Only the three known *Gli* genes were detected demonstrating that the mouse genome is unlikely to contain additional *Gli* genes. In addition YACs containing the nearest flanking marker to *Hx* were isolated and were shown not to contain a *Gli* gene.

The relationship between *Hx* and a second gene, *Pmsc2*, was also investigated. *Pmsc2* is a subunit of the 26S proteasome, a multi-subunit complex performing the degradation of polyubiquitinated proteins. The developmental expression pattern and genetic map position of *Pmsc2* implicated *Pmsc2* as a candidate for *Hx*. This was assessed by defining the genetic position of *Pmsc2* relative to the closest flanking markers to *Hx*. Analysis of the European Collaborative Interspecific backcross, EUCIB, demonstrated *Pmsc2* to be located 4.4cM proximal to *En2*, which is itself proximal to *Hx*. *Pmsc2* is therefore excluded as a candidate for *Hx*.

# TABLE OF CONTENTS

**Chapter 4 Mutational analysis of *Gli2* and *Inhβb* in *Dh* animals**



**Chapter 5 Attempted identification of a novel *Gli* family gene, a potential candidate for *Hemimelic extra toes***

# LIST OF FIGURES

# LIST OF TABLES

## ACKNOWLEDGEMENTS

I would like to express my gratitude and appreciation to my supervisor, Dr. Bob Hill, for his excellent guidance, support and encouragement. I am indebted to him and indeed to the whole Human Genetics Unit for providing such a stimulating and congenial environment in which to learn my trade. Big thanks also to David Hughes for answers to endless questions, a steady stream of expansive ideas and, not least, for instilling in me the importance of combating that perpetual battle against spiralling entropy.

Thanks are also due to all the many people who helped me throughout my time at the H.G.U. Special thanks to Muriel Lee for excellent and beautiful FISH data, Andreas Schedl for his YAC expertise, Cathy Abbott for invaluable mapping advice, John West for more invaluable mapping advice, Gordon Allan and Laura Lettice for help with chugging through endless sequencing gels, Agnes for oligos galore and Douglas Stuart for his invaluable figure talents.

A huge thanks to Jane for her deep love and support .....and then, again. A equally huge thanks to my good buddies Dave, Jack and Rich. I now know more about what friendship really is. You helped me more than you know.

Of course I wouldn't be here without the love and exceptional upbringing given to me by my mother and father. Thank you for everything.

# CHAPTER 1

# INTRODUCTION

## 1.1 Vertebrate limb development: a model system for studying the processes of vertebrate morphogenesis

The developing limb is one of the foremost systems in which to study the mechanisms which underlie vertebrate morphogenesis. Historically the limb has long been a favourite organ of developmental biologists due mainly to its external location and hence easy accessibility. Experimental analysis of limb development has largely employed chick embryos because they are amenable to experimental manipulations *in ovo*. Traditionally the effects of such manipulations have been observed by whole mount staining procedures, histology and morphological examination. This has provided extensive embryological data concerning the cellular interactions within the limb bud, (reviewed by Carlson, 1988) and has enabled the identification of the two major signalling centres, the apical ectodermal ridge (AER) and the zone of polarising activity (ZPA), which mediate limb bud outgrowth and patterning across the anterior-posterior axis respectively. As a result of work carried out in both chick and mouse a growing number of molecular markers have been identified which allow a detailed picture to be assembled of how particular experimental perturbations influence molecular and cellular processes in the limb.

The limb is a complex organ, the development of which involves fundamental processes common to development throughout the embryo. Mechanisms such as pattern formation, induction, epithelial-mesenchymal interaction, programmed cell death, connective tissue, muscle and bone formation, angiogenesis and innervation are all intrinsic to normal limb development. Experimentation leading to insight into these processes will therefore further the understanding not only of limb development but also of more inaccessible and intractable areas and organs. Consistent with this is that many of the mouse limb mutations currently described (of which there are over 60)

are pleiotrophic (Green, 1989). This implies that these genes either function in multiple molecular pathways which operate specifically in different embryonic areas or, more likely, that similar molecular pathways are involved in the development of multiple organs. The latter idea is beginning to emerge from the 'bottom up', in that genes which form the molecular basis of limb morphogenesis are also intimately involved in the development of many other regions, from the initial establishment of the body plan to the patterning of organ systems (see Cohn and Tickle 1996, and refs. therein).

A further reason to study limb development is to shed light on congenital malformations. Human limb dysmorphologies usually fall into one of three categories: (1) reduction defects, in which a part or an entire limb is missing (hemimelia and amelia); (2) duplication defects, in which supernumerary digits arise (polydactyly); (3) loss of digits (oligodactyly); and (4) dysplasia of the limb in which fusion of the bones occurs, particularly of the digits (syndactyly). Additionally, these defects may be associated with craniofacial abnormalities or dysgenesis of internal organs, analogous to the pleiotrophic nature of the mouse limb mutations. Polydactyly is the most frequent human congenital hand deformity (McCusick, 1992), and has attracted attention at least since biblical times; "*And there was yet a battle in Gath, where was a man of great stature, that had on every hand six fingers, and on every foot six toes, four and twenty in number; and he also was born to the Giant.*" (Second book of Samuel, chapter 21; 20, The Holy Bible). These disorders, affecting digit formation, have been classified according to morphology (Winter and Tickle, 1993). Recent and future discoveries will enable a classification based on genetic determinants to be made.

Animal models will be important for understanding these syndromes, however to date only one mouse model has been identified which shares the genetic defect of a human limb deformity syndrome. Greig cephalopolysyndactyly syndrome presents polysyndactyly and craniofacial abnormalities and is due to haploinsufficiency caused by lesions in *GLI3*, a zinc finger gene. Correspondingly the

mouse mutant *Extra toes*, *Xt*, is also due to haploinsufficiency of the *Gli3* locus (this is discussed in detail in section, 1.3.6, and in chapter 4, 4.1.2). The genetic basis for several other human syndromes have recently been determined, although no corresponding animal models have been identified. For example synpolydactyly (or Type II syndactyly), characterised by altered growth and branching patterns in the distal skeletal elements of both hands and feet, is caused by mutation of the *HOXD13* gene (Muragaki *et al.*, 1996). Also a number of related disorders have been shown to be due to mutations in the fibroblast growth factor receptor (*FGFR*) family: Crouzon syndrome, *FGFR2*; Pfeiffer syndrome, *FGFR1* and *FGFR2*; Apert syndrome, *FGFR2*; Jackson-Weiss syndrome, *FGFR2*; and achondroplasia, *FGFR3* (reviewed by Muenke and Schell, 1995). Further understanding of these conditions will be possible by the generation of these specific mutations in mice through the technology of gene targeting (Joyner, 1993).

Limb morphogenesis requires the concerted action of several hundred genes, however many remain to be discovered. To further understand limb development therefore, the work presented in this thesis concerns the identification of the genes responsible for two mouse mutations; *Dominant hemimelia*, *Dh*, and *Hemimelic extra toes*, *Hx*, both of which display a limb phenotype. This chapter gives an introduction to limb development, the *Dh* and *Hx* mutations and the methodologies of mouse molecular genetics which enable the cloning and functional analysis of such genes. Introductions are presented at the beginning of each results chapter concerning the relevant biology and the technologies employed.

## 1.2 Vertebrate limb development

### 1.2.1 Overview of the early events in limb development

In virtually all vertebrates, two pairs of limbs develop at defined axial levels from small buds that consist of mesenchymal cells encased in an ectodermal hull. The initial formation of the limb bud is caused by a condensation of cells from the lateral plate mesoderm, accompanied by a selective decrease in proliferation of cells on either side of the prospective buds (Searls and Janners, 1971). Once formed the bud grows out from the body wall due to mesenchymal proliferation at the tip of the bud, while mesenchymal cells at the base of the bud, below the region of proliferation, begin the process of differentiation into the various tissues of the mature limb. Subsequently the tip of the bud flattens to produce the hand/foot plate wherein the digits form (Carlson, 1988, presents a detailed review of limb development at the morphological level). It is useful to specify the axes of the limb in a Cartesian manner. The proximodistal axis goes from the shoulder to the tips of the fingers; the anterioposterior axis from the little finger to the thumb (which is also termed postaxial to preaxial); and the dorsoventral axis which runs from the top of the hand to the palm. Experimental analysis, principally embryological manipulations performed *in ovo* on chick limbs, has uncovered discrete cell populations which mediate pattern formation and mophogenesis within the limb.

The apical ectodermal ridge (AER) arises from surface ectoderm just after the limb primordium projects from the flank. It forms a rim of pseudostratified, elongated epithelial cells at the apex of the bud. The cells are linked by extensive gap junctions and are tightly packed, providing a degree of rigidity to the bud (Fallon and Kelley, 1977). The AER is induced by the underlying mesenchyme (Kieny, 1968) and interactions between the AER and an area of mesenchyme directly beneath the AER, known as the progress zone, are continuous and reciprocal throughout early development. Saunders (1948) was the first to show this critical relationship. When the AER is removed the bud ceases to grow and a truncated limb results. Timing of

AER removal determines the degree to which the limb will be truncated. Early removal results in the failure of proximal structures to form, while removal at a later time allows proximal structures to develop normally but results in truncation of the most distal elements (Saunders, 1948; Summerbell, 1974). This observation reflects the sequential formation of structures along the proximodistal axis, with proximal elements being determined before more distal ones. Although the AER is essential for bud outgrowth recombination experiments show that it does not provide the signals which specify pattern within the limb mesoderm (Rubin and Saunders, 1972; Zwilling, 1956).

The progress zone consists of undifferentiated mesenchymal cells which have a high division rate and are responsible for the growth of the bud. As division occurs cells leave the progress zone and begin to differentiate. Truncation caused by AER removal occurs due to the cessation of proliferation by these cells. Transplantation studies show that the progress zone behaves autonomously. Grafting the progress zone of an early bud to an older bud results in a duplication of proximodistal elements. Conversely, replacement of the progress zone from an early bud with that of an older bud results in the absence of proximal structures. (Summerbell *et al.*, 1973). Thus the cells of the progress zone appear to be subject to a timing mechanism, the length of time a cell resides within the progress zone determines identity along the proximodistal axis.

In 1968 Saunders and Gasseling discovered a major signalling centre for the specification of pattern across the anterioposterior axis. Grafting a region of cells from the posterior bud margin to the anterior margin of a host chick wing results in the development of supernumerary digits in a reverse orientation with respect to the normal pattern; instead of the normal 234 pattern (anterior to posterior) a 432234 pattern is produced. The induced digits are derived from host cells therefore this region, termed the zone of polarizing activity (ZPA) is a true signalling centre. The cells which constitute the ZPA are not discernible histologically and were mapped in grafting experiments by the ability to induce supernumerary digits (Tickle, 1981;

Honig and Summerbell, 1985; Hinchliffe and Sansom, 1985). The highest degree of polarizing activity is located posteriorly, near the tip of the bud just proximal to the progress zone and is confined to the mesenchyme.

The dorsoventral polarity of the limb is established before the emergence of the limb bud. Saunders and Ruess (1974) showed that grafting limb field mesenchyme in an opposite dorsoventral orientation led to limbs with reversed dorsoventral polarity. Studies then showed the cues specifying dorsoventral polarity are restricted to the ectoderm. Grafting the ectodermal jacket in an inverted dorsoventral orientation onto the mesoderm of a host bud gives rise to limbs which are also partially inverted dorsoventrally (MacCabe *et al.*, 1974). A greater degree of inversion is observed proximally suggesting that only the cells of the progress zone are competent to interpret dorsoventral cues from the ectoderm. It is likely that signals arise both dorsally and ventrally because ectopic limbs induced by grafting an AER to either the dorsal or ventral surface of a chick wing bud display only a dorsal or ventral configuration respectively (Saunders *et al.*, 1976; Shellswell and Wolpert, 1977).

### 1.2.2 Limb bud initiation

All vertebrates are limited to a maximum of four limbs, however very little is known about the factors which specify limb induction, particularly limb number, type and position. In different vertebrates limbs arise opposite different somite levels, however their position appears consistent with respect to patterns of homeobox gene expression along the body axis (Burke *et al.*, 1995).

In the chick ectopic limbs can be induced by the application of a bead soaked in FGF (FGF1, -2, -4, or -8) to the presumptive flank, which lies between wing and leg (Cohn *et al.*, 1995; Crossley *et al.*, 1996). The entire flank has limb forming potential with anterior flank giving additional forelimbs and posterior flank giving additional hindlimbs. Intriguingly these ectopic limbs have a reversed anterioposterior polarity (the normal digit pattern from anterior to posterior is 234, the reverse

pattern being 432). This reverse pattern can be explained by potential polarising activity possessed by flank cells, as revealed by grafting these cells to the anterior margin of a wing bud (Hornbruch and Wolpert, 1991). This polarising activity is graded with the highest polarising potential located anteriorly. Thus induction of an ectopic limb will recruit cells with the highest polarising activity anteriorly rather than posteriorly as in the normal limb. One consequence of FGF application on gene expression is the activation of *Fgf8* in the body wall ectoderm which is soon followed by activation of *Sonic hedgehog* (*Shh*) resulting in a polarizing region. During normal limb development *Fgf8* transcripts are detected in the intermediate mesoderm near the presumptive limb (at the level of the wing in the nephrogenic mesoderm cells adjacent to the Wolffian duct and at the hindlimb level in the primitive streak) at the correct time for limb induction (Crossley *et al.*, 1996). It is thus likely that an endogenous FGF signal, perhaps FGF8, could initiate limb bud development but it is presently unclear. It is likely that differential competence of the lateral plate mesoderm to respond to a limb inducing signal also plays a role in determining the site of limb formation. Transplantation studies provide direct evidence that prospective interlimb region mesoderm is less competent to respond to the endogenous inducer than prospective wing territory mesoderm (Geduspan and Solursh, 1992).

Recently Cohn *et al.* (1997) report the anterior boundaries of *Hox* group 9 paralogues, *Hoxb9*, *c9* and *d9* in the lateral plate mesoderm and show that local application of FGF to the flank resets these expression boundaries. FGF applied to the anterior flank, resulting in an ectopic wing, transformed the normal flank pattern to a pattern normally found at wing level, while FGF applied to the posterior flank, resulting in an ectopic leg, transformed the normal flank pattern to a pattern normally found at leg level. The fact that an additional limb forms, rather than a shift in position of the nearby normal limb, suggests that cells are already committed to a specific limb position. Rather FGF may interfere with regulatory genes which maintain the pattern of *Hox* expression. This data argues that the early primary

expression of *Hox* genes in lateral plate mesoderm specifies limb position. Consistent with this is the *Hoxb5* mutant in which the level of the forelimb is shifted (Rancourt *et al.*, 1995).

Michaud *et al.*, (1997) have performed recombination experiments using chick-quail chimeras and suggest a fate map for the origin of wing bud mesoderm and ectoderm. Prospective wing mesoderm is localised in the medial part of the somatopleure (at stage 13), while the overlying ectoderm is entirely fated to become AER. Juxtaposition of two rows of somites with the prospective wing territory induced bidorsal limbs but with normal AERs, while insertion of a barrier between the prospective limb field and lateral somatopleure mesenchyme also produced bidorsal limb buds, again with a fully developed AER. These studies indicate that the dorsoventral polarization of the limb field mesenchyme is achieved through opposition of dorsalising and ventralising signals released by the somites and the lateral somatopleure respectively. These signals would then act to polarise overlying ectoderm (between stages 13 to 15). Failure to disrupt the formation of a normal AER however indicates that AER induction occurs independently of dorsoventral polarisation of flank ectoderm by the underlying mesenchyme.

### 1.2.3 Molecular basis of pattern specification in the early limb bud

The concept of pattern formation requires that a four dimensional coordinate system is initiated such that cells are instructed as to their positional identity in space and time. This information must then be interpreted to direct the generation of the appropriate structures (Wolpert, 1971). Recently molecules have been identified which operate to specify patterning along each axis of the limb bud. Also mutational analysis has putatively revealed *Hox* genes as being involved in controlling limb patterning by regulating the timing of the developmental program.

Signalling by the AER is required for the proximodistal outgrowth of the limb bud and there is substantial evidence to suggest that this is mediated by FGFs. *Fgf4*

is detected in early mouse buds in a restricted pattern from the apex to the posterior boundary of the AER (Niswander and Martin, 1992), while FGF2 immunoreactivity is detected throughout the AER, in a thin layer of subectodermal mesenchyme and in the dorsal surface ectoderm (Savage *et al.*, 1993). Removal of the AER from early chick limb buds leads to truncation but outgrowth and patterning can be maintained by stapling FGF2 or FGF4 soaked beads in place of the ridge. *Fgf8* however is currently the best candidate as the endogenous AER signal that mediates bud outgrowth. The induction of *Fgf8* expression in the surface ectoderm precedes the first mophological manifestation of limb bud outgrowth and expression continues in the AER throughout limb bud outgrowth (Heikinheimo *et al.*, 1994; Crossley and Martin 1995). During limb bud induction FGF8 can stimulate outgrowth of the lateral plate mesoderm and FGF8 soaked beads are also able to maintain outgrowth and patterning in limb buds following AER removal (Crossley *et al.*, 1996).

The progress zone, directly underlying the AER, must be competent to respond to signals from the AER and maintain an undifferentiated state. Genes expressed in the progress zone include *Msx1* and *Msx2*; *Msx1* is strongly expressed in the mesenchyme and also weakly in the AER while *Msx2* is more weakly expressed in the mesenchyme and more strongly in the ridge (Hill *et al.*, 1989; Robert *et al.*, 1989). Tissue transplantation studies and ridge removal experiments show that expression of these genes is highly regulated by AER signals (Davidson *et al.*, 1991; Robert *et al.*, 1991) demonstrating the interplay between AER and the underlying mesenchyme. Transfection of *Msx1* prevented muscle differentiation in a myogeneic cell line indicating a role in maintaining the progress zone in an undifferentiated state (Song *et al.*, 1992).

Two molecules have been shown to mediate the polarizing activity of the ZPA. Retinoic acid (RA) was the first molecule identified which could recapitulate the ability of anteriorly grafted ZPA to duplicate wing patterns (Tickle *et al.*, 1982;

Tickle *et al.*, 1985; Summerbell *et al.*, 1983). This effect and the role of RA in relation to the endogenous mechanisms which determine the ZPA remains unclear, however several observations support the role of RA in normal limb development. Firstly RA is enriched in the ZPA with the concentration of RA required to induce digit formation being the same as the endogenous RA concentration. Secondly limb buds express all the known components of retinoid signalling; receptors, binding proteins and metabolic enzymes. Thirdly RA induces the expression of posterior specific genes such as *Hoxd10*, *-11*, *-12* and *-13* (reviewed by Tickle and Eichele, 1994).

Cells from a region immediately surrounding an RA releasing bead acquire ZPA activity indicating that RA induces the signals required for ZPA activity rather than RA acting as a morphogen itself. Indeed RA induces expression of *Shh*, the second molecule known to confer ZPA activity (Riddle *et al.*, 1993). *Shh* is an orthologue of the *Drosophila* segment polarity gene *hedgehog* (*hh*) and has attracted much attention as it is expressed in many organising centres during embryogenesis (reviewed by Hammerschmidt *et al.*, 1997). The expression of *Shh* in the limb correlates tightly with the ZPA and misexpression studies at the anterior of the limb bud show *Shh* to induce mirror image duplication of distal skeletal elements (Riddle *et al.*, 1993; Chang *et al.*, 1994).

*Shh* is a secreted signalling molecule and a major question about *Shh* signalling concerns its range; the polarizing signal is required to relay information 100-150μm, a considerable distance. If *Shh* acts as a morphogen then it should be distributed in a gradient across the anterioposterior axis, with high concentrations of SHH specifying posterior elements and lower concentrations giving progressively more anterior identities, however the pattern of SHH across the limb bud has not been reported. *Shh* may act in an analogous manner to *hh* in *Drosophila* wing imaginal discs, exerting a long range effect by regulating the expression of some other signalling molecule; *hh* activates *decapentaplegic* (*dpp*), a *Tgfβ* family member (Basler and Struhl, 1994). A candidate for such a mediator is the bone morphogenetic protein, *Bmp2*, which is expressed in a domain slightly larger than that of *Shh* (Laufer *et al.*, 1994; Francis *et*

*al.*, 1994). Further *Bmp2* is activated in response to *Shh* and *Fgf4* (see below). BMP2 however exerts no polarizing activity (Francis *et al.*, 1994) but it may be that BMP2 acts together with some other factor, perhaps as a heterodimer with another BMP. *Bmp4* and *Bmp7* are also expressed in the mesenchyme and AER (Francis *et al.*, 1994).

Genetic studies in *Drosophila* have identified several components of the signal transduction pathway which mediates the *hh* signal (reviewed by Hammerschmidt *et al.*, 1997). *hh* acts by binding to a cell surface receptor, the nature of which is complex. Candidate receptors include *patched* (*ptc*) which encodes a transmembrane protein with a structure similar to that of channels and transporters (Hooper and Scott, 1989; Nakano *et al.*, 1989) and *smoothened* (*smo*) (van den Heuvel and Ingham, 1996; Alcedo *et al.*, 1996), a seven transmembrane protein with structural similarities to G-protein coupled receptors. Recent biochemical evidence suggests that PTC binds to SMO and SHH, whereas SHH does not bind to SMO (Stone *et al.*, 1996; Marigo *et al.*, 1996). One model suggests that *smo* is a constitutive activator of *hh* target genes. This activity is normally repressed by *ptc*, but this repression is relieved by *hh* binding to *ptc*. Five genes have been identified which encode cytoplasmic components of the pathway; *fused* (*fu*), specifying a putative protein serine-threonine kinase for which no substrates have yet been found, *suppressor of fused* [*Su(fu)*] whose molecular nature is unknown, *costal2* (*cos2*) a kinesin-like motor protein, *pka* which encodes a cAMP-dependant protein kinase A and *cubitus interruptus* (*ci*) a zinc finger protein (see Hammerschmidt *et al.*, 1997 for refs.).

A good candidate for the point of convergence of this pathway is *ci*, a putative direct transcriptional activator of known *hh* target genes and which is down regulated post-translationally by *pka* and *ptc*, both of which negatively regulate *hh* signalling (Johnson *et al.*, 1995; Dominguez *et al.*, 1996). Recent findings suggest that CI resides within a complex involving FU and COS2 which is bound with high affinity to microtubules in the cytoplasm. Stimulation of cells by HH releases the

complexes from the microtubules and facilitates the release of CI from the complex. This suggests a means by which CI is sequestered in the cytoplasm but in response to HH CI is released to facilitate transfer to the nucleus (Robbins *et al.*, 1997; Sisson *et al.*, 1997). Aza-blanc *et al*, (1997) provide evidence that this release is mediated by proteolysis to release a transcriptionally active form of CI from a tethering domain.

Growing evidence indicates that the hedgehog signalling pathway, or at least components thereof, are conserved from flies to mammals. Vertebrate homologues have been identified for many of the genes involved in hedgehog signalling including: *Shh*, *Desert hedgehog* (*Dhh*), *Indian hedgehog* (*Ihh*), orthologues of *hh*; *Gli*, *Gli2* and *Gli3*, orthologues of *ci*; *Ptc*, orthologue of *ptc*; *Pka*, orthologue of *pka*; *Bmp2*, *Bmp4* and *Bmp7*, orthologues of *dpp*, *Wnt* genes, orthologues of *wingless* (*wg*) and *En1* and *En2*, orthologues of *engrailed* (*en*).

The *Gli* orthologue *ci* is an important component of the *hh* pathway. It is required for the activation of *hh* target genes such as *wg*, *dpp* and *ptc* (Motzny and Holmgren, 1995; Johnson *et al.*, 1995; Dominguez *et al.*, 1996). *ci* can also function as a repressor of *hh* expression (Dominguez *et al.*, 1996). In an analogous fashion mutation of mouse *Gli2* and *Gli3* results in defective *Shh* and *Ihh* signalling and causes ectopic expression of *Shh*, leading to digit duplication (Mo *et al.*, 1997). As predicted by *Drosophila* observations *Bmp2*, *Bmp4* and *Ptc* are expressed in regions complementary to or overlapping with *hedgehog* expressing cells and can be induced by ectopic *Shh* expression (Goodrich *et al.*, 1996; Margio *et al.*, 1996). In addition the role of *Pka* as a negative regulator of the *hh* pathway is conserved in *Drosophila* and vertebrates (Fan *et al.*, 1995; Hynes *et al.*, 1995; Hammerschmidt *et al.*, 1996).

Several human diseases result from mutation in *hedgehog* signalling genes. Point mutations in *SHH* itself have been shown to be responsible for some cases of familial holoprosencephaly (Belloni *et al.*, 1996; Roessler *et al.*, 1996). *PTC* appears to play a role as a tumour suppressor gene, mutations being found to underlie the autosomal disorder Nevoid basal cell carcinoma syndrome (NBCCS) which is characterised by developmental abnormalities and a predisposition to various forms

of cancer including basal cell carcinomas (Johnson *et al.*, 1996; Hahn *et al.*, 1996). In addition mutations in *PTC* were found in sporadic cases of basal cell carcinomas (Johnson *et al.*, 1996; Hahn *et al.*, 1996). Mutations in *GLI3* causing Grieg cephalopolysyndactyly syndrome (GCPS) and Pallister-Hall syndrome (PHS) are discussed below (1.3.5 and in chapter 4) while *GLI* is implicated in the cancers glioblastoma and childhood sarcoma (see chapter 4).

*Shh* and *Fgf* signals activate a number of *Hox* genes in the developing limb which show complex dynamic patterns of expression (reviewed by Izpisua-Belmonte and Douboule, 1992). *Hoxa10*, *-11* and *-13* expression is initiated posterodistally in the mesenchyme and subsequently expands anteriorly so that the final boundaries of expression are perpendicular to the proximodistal axis. The boundaries of expression are nested with the *Hoxa10* the most proximal and that of *Hoxa13* the most distal. 5′ *Hoxd* genes, *Hoxd9*, *-10*, *-11*, *-12* and *-13* also show a nested pattern of expression. *Hoxd9* and *-10* are expressed throughout the mesenchyme, while *Hoxd11*, *-12* and *13* are progressively more restricted posteriorly. As the limb bud grows out these patterns of expression appear to be pulled out with the extending tip, except for *Hoxd13* which remains confined to the tip of the bud. By the stage when digits are differentiating *Hoxd12* and *-13* have extended more anteriorly, with *Hoxd13* being expressed throughout the mesenchyme at the tip of the bud.

Polarizing signals, in the presence of an AER or FGF4, produce mirror image duplication of *Hoxd* expression patterns in a manner which correlates with digit duplication (Izpisua-Belmonte *et al.*, 1992). This, in combination with the multifarious *Hoxa* and *-d* expression patterns has lead to the suggestion that *Hox* genes encode positional information. However gene disruption of *Hoxd11*, *-13* and *Hoxa11* does not produce malformations consistent with a role in simply determining position. Rather this data is consistent with a role in the regulation of timing and the extent of local growth rates (Dolle *et al.*, 1993; Davis and Capecchi, 1994; Favier *et al.*, 1995; Davis *et al.*, 1995).

Recent studies have begun to illuminate molecular signals involved in dorsoventral patterning. The *Drosophila* gene *fringe* is characterised as a putative secreted factor controlling wing margin formation. The formation of a wing margin, which subsequently organises outgrowth, has been shown to be formed by the juxtaposition of *fringe* expressing cells and non-expressing cells in the wing imaginal disc (Irvine and Wieschaus, 1994; Kim *et al.*, 1995). The vertebrate orthologue *Radical fringe* (*R-fng*) is also a secreted molecule which is expressed in the presumptive dorsal limb bud ectoderm and remains restricted there during outgrowth, with highest expression in the AER. Localised, ectopic expression of *R-fng* in ventral ectoderm leads to the induction of additional, ventrally located AERs, whereas uniform expression partially or completely suppresses AER formation. This suggests that AER positioning and induction is regulated by the juxtaposition of *R-fng* expressing and non-expressing cells, in an analogous manner to the formation of a wing margin in *Drosophila* (Laufer *et al.*, 1997; Rodriguez-Esteban *et al.*, 1997).

*Wnt7a* is a secreted signalling molecule expressed in presumptive dorsal limb flank prior to limb bud protrusion and later remains restricted to the dorsal limb bud ectoderm (Dealey *et al.*, 1993; Parr *et al.*, 1993). Evidence for a role in dorsoventral polarity comes from the inactivation of *Wnt7a* which causes ventralisation of the paws, although the stylopod and zeugopod remain unaffected (Parr and McMahon, 1995). A mesenchymal target for *Wnt7a* appears to be the *Lim* homeodomain gene *Lmx1*. *Lmx1* is activated by *Wnt7a* in chick limb dorsal mesenchyme (Riddle *et al.*, 1995), while removal of dorsal limb bud ectoderm down regulates *Lmx1*. Furthermore ectopic *Lmx1* induces ventral to dorsal transformations of limb structures (Vogel *et al.*, 1995).

*Engrailed-1* (*En1*) is a homeobox containing gene which mediates ventral specification of the limb bud ectoderm. *En1* is expressed in the ventral ectoderm up to the midline of the AER (Loomis *et al.*, 1996). Inactivation of *En1* in the mouse results in dorsalisation of both ectodermal and mesenchymal limb structures and causes ectopic expression ventrally of *Wnt7a*. In addition the AERs of *En1* knockout

mice are expanded ventrally and display an abnormal, flattened morphology (Loomis *et al.*, 1996). Ectopic expression of *En1* in chick limb buds represses expression of both *Wnt7a* (Logan *et al.*, 1997) and *R-fng* (Laufer., *et al.*, 1997) and leads to disruption of AER formation. These data suggest that *En1* antagonises dorsal fate, through *Wnt7a* and, and restricts the ventral AER border through antagonism of *R-fng*.

### 1.2.4 Coordination of signalling networks

Traditionally the signalling mechanisms operating in the three main axes have been viewed as acting independently. However with the functional investigation of the genes discussed above it has become apparent that pattern formation is intimately interconnected and coordinated across the three axes.

Niswander *et al.*, (1994) and Laufer *et al.*, (1994) were the first to show that signalling from the mesenchyme and the AER is mutually controlled. Using a combination of surgical manipulations and applications of ectopic SHH (or retinoic acid) and FGF4 a positive feedback loop was established at the epithelial/mesenchymal junction. Removal of posterior AER greatly reduces the normal expression of *Shh* in the posterior margin of the bud (and also ablates polarizing activity of the posterior mesoderm, Vogel and Tickle, 1993; Niswander *et al.*, 1993). However application of the ridge signal FGF4 to the posterior tip of the limb, following AER removal, recovers normal *Shh* levels. Therefore maintenance of *Shh* may require FGF4.

Application of ectopic *Shh* to the anterior margin of the limb bud induces an anteriorly extended AER in which an ectopic domain of *Fgf4* expression is induced (*Fgf4* is normally restricted posteriorly in the AER) and anterior mesenchyme proliferates. In addition ectopic expression of *Hoxd9*, *Hoxd13* and *Bmp2* are induced in a mirror image to endogenous expression. Application of ectopic *Shh* to the anterior margin of the limb bud after removal of the anterior AER however induces no cellular proliferation or ectopic *Hoxd9*, *Hoxd13*, *Bmp2* expression. However when

*Shh* is applied to the anterior margin of the limb bud after removal of the anterior AER in combination with FGF4 at the tip of the anterior bud cellular proliferation and ectopic *Hoxd9*, *Hoxd13*, *Bmp2* expression are reproduced. Application of FGF4 in the absence of *Shh* does not induce *Hoxd* or *Bmp2* expression but FGF4 alone induces mesodermal outgrowth. Therefore *Fgf4* expression in the AER can be activated and sustained by *Shh* and mesodermal competence to respond to *Shh* can be mediated by FGF4. Further mesoderm proliferation in response to *Shh* must be indirect, being mediated by the induction of FGFs. Only *Fgf4* has been investigated in these experiments but it is possible that *Fgf4* may be mimicking the activity of another *Fgf* family member, such as *Fgf8* or *Fgf2* or an as yet unknown member.

The induction of *Fgf4* expression by *Shh* in the ectopic AER and the maintenance of *Shh* expression by *Fgf4* suggest that *Shh* and *Fgf4* expression are normally sustained by a positive feedback loop. This coordination is possible because *Shh* patterns mesodermal tissue and regulates *Fgf4* expression, while *Fgf4* induces mesodermal proliferation and maintains *Shh* expression and also confers a competence to the mesoderm to respond to *Shh*. Such a mechanism provides a strict coordination of patterning and proliferation. This model is supported by *Shh* mouse mutants which lack distal limb structures, presumably as a secondary consequence of the loss of *Fgf* expression in the AER (Chiang *et al.*, 1996).

Determination of the dorsoventral axis has also been shown to interact with components of anterioposterior patterning. In the *Wnt7a* knockout mouse there is a notable loss of posterior limb skeletal elements; digit five and other digital elements and the ulna are abnormal and are often missing. This absence is due to a reduction in the size of the *Shh* expression domain leading to an attenuation of the ZPA signal (Parr and McMahon, 1995). Furthermore removal of dorsal ectoderm from chick wing buds results in a reduction of *Shh* expression, however this can be prevented by grafts of cells expressing *Wnt7a* (Yang *et al.*, 1995).

Posterior restriction of the signalling loop may be controlled by the restricted expression of a *Hox* gene. *Hoxb8* is normally expressed posteriorly along the body

axis and in posterior forelimb bud mesenchyme. Under the control of the retinoic acid receptor β2 promotor, Charitie *et al*, (1994) ectopically expressed *Hoxb8* in the anterior mesenchyme of transgenic mice. This lead to the induction of an ectopic signalling loop between *Shh* and *Fgf4* in the anterior of the forelimb bud and mirror image duplication of digits. It is tempting to speculate that a second *Hox* gene establishes the position of an identical *Shh*, *Fgf*, *Wnt7a* signalling loop in the posterior of the hindlimb bud.

The intermediates in the signalling loops between *Shh*, *Fgf*s and *Wnt7a* have not been defined. Members of the *Bmp* family may however interpose between *Shh* and *Fgf*s. In the chick mutant *talpid*, which develop supernumerary, morphologically similar digits, *Shh* expression remains posteriorly restricted but *Bmp2*, *-4* and *-7* are expressed ectopically, in a uniform pattern across the mesenchyme while *Fgf4* is expressed throughout the AER. *Bmp* genes may therefore mediate between *Shh* and *Fgf4* (Francis-West *et al*., 1995).

### 1.3.1 luxoid group of limb mutations

Despite the recent advances in identifying the molecular pathways underlying limb development much still remains unknown. For example how is limb position and identity, either fore or hind, determined? What is the cascade of events which initiates the inductive properties of the AER and ZPA? How is the information conferred by the different signalling molecules integrated and what are the intermediate steps and target genes of the signal transduction pathways? Clearly investigation of these and other important questions are currently inhibited by the fact that many genes which are fundamentally involved have not yet been identified. This point is well illustrated by the large catalogue of mouse mutations displaying a limb phenotype for which the genes involved are not known. At present there are over 60 such mutants which together present defects affecting every aspect of limb morphogenesis (Green, 1989). As such they represent a valuable resource for the identification and functional study of important limb genes. The recent advances in mouse and molecular genetic techniques which enable such investigation are discussed below (1.4).

Of the many limb mutations, Searle (1964) proposed a distinct group, the 'luxoid' group, on the basis of phenotypic similarity in the aberrant limb. The characteristic phenotype is a twisted appearance of the forelimbs, hindlimbs or both. This is caused by luxation; dislocation and twisting of the long bones, and hemimelia; partial or complete loss of one or more of the long bones. Six mutations were originally placed in the luxoid group; *Dominant hemimelia* (*Dh*), *Luxoid* (*Lu*), *Luxate* (*Lx*), *Strong's luxoid* (*Lst*), *oligodactyly* (*ol*) and *postaxial hemimelia* (*px*). Subsequently *Extra toes* (*Xt*) (Johnson, 1967) and *Hemimelic extra toes* (*Hx*) (Dickie, 1968) were described which reasonably fit into the luxoid group. Additional defects include digital abnormalities; increased numbers of digits (polydactyly), decreased numbers of digits (oligodactyly) or fusion of the digits (syndactyly). Apart from the limb phenotype certain members of the group also present limited axial skeleton and visceral defects but these abnormalities are variable between group members. The

phenotypic characteristics of each mutant is presented in table 1.1. All the mutants are semidominant, except for *ol* and *px* which are recessive. Penetrance of the limb phenotype is almost complete, typically 95%, but variable in expressivity.

A common feature of the limb phenotypes is the non-equivalence of the developing long bones. The majority of the luxoid group *Dh*, *Lu*, *Lx*, *Hx*, *Xt* and *Lst* cause preaxial hemimelia of the tibia while the fibula remains normal (except for occasional bowing, presumably due to a shortened or absent tibia). Also digit defects for these mutants are preaxial, the only exception being an occasional postaxial nubbin observed in *Xt*. The converse is true for *px* and *ol* which present postaxial hemimelia of the ulna and fibula with postaxial digit defects. This differential effect may indicate an intrinsic aspect of the patterning process. A possible explanation is that during normal development the preaxial prechondrogenic condensations are formed after those of the postaxial side. This suggests therefore a possible defect in developmental timing or growth rates of the condensing mesenchyme.

In addition to the phenotypic similarities a genetic relationship has been shown between members of the group. Crosses to generate the compound mutants containing either *Lx* and *Dh* (Searle, 1964) or *Lx* and *Lu* (Forsthoefel, 1959) demonstrate genetic interaction. The phenotype of *Dh*/+*Lx*/+ animals shows an enhancement of the *Dh*/+ phenotype with increased oligodactyly and more severe hemimelia. In *Dh*/+*Lx*/*Lx* animals the stomach is severely reduced, again an enhancement of the *Dh*/+ phenotype. Compound mutants, carrying all combinations of heterozygosity or homozygosity of the *Lx* and *Lu* alleles display a consistent enhancement of the phenotypes seen for the individual alleles. The severity of phenotype correlates with the mutation dosage. For example the limb phenotype of *Lx*/+*Lu*/+ animals is very similar to that of either *Lx* or *Lu* homozygotes.

Two explanations can be proposed for the common limb phenotype and genetic interaction demonstrated for the luxoid group. Firstly they may be due to mutation in a family of homologous genes which play similar developmental roles. Identification of one luxoid gene may therefore facilitate the identification of other

**Table 1.1** Phenotypic characteristics of the luxoid mutants. ++ = heavily affected, + = affected, (+) = slight or occasional affect, - = not affected. With presacral vertebrae, d = number decreased, I = number increased. (The *Hx* homozygote phenotype has not been reported).

| Structure affected or nature of abnormality | *Dominant hemimelia* | | *Extra toes* | | *Hemimelic extra toes* | *Luxate* | | *Luxoid* | | *Strongs' luxoid* | | *oligodactyly* | *postaxial hemimelia* |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | *Dh/+* | *Dh/Dh* | *Xt/+* | *Xt/Xt* | *Hx/+* | *lx/+* | *lx/lx* | *Lu/+* | *Lu/Lu* | *Lst/+* | *Lst/Lst* | *ol/ol* | *px/px* |
| Forelimbs | - | - | + | + | + | - | - | - | + | (+) | + | + | + |
| Hindlimbs | + | + | + | + | + | + | + | + | + | + | + | + | + |
| Preaxial side | + | + | + | + | + | + | + | + | + | + | + | - | - |
| Postaxial side | - | - | (+) | (+) | - | - | - | - | - | - | - | + | + |
| Polydactyly | + | - | + | ++ | + | + | + | + | + | + | ++ | - | - |
| Oligodactyly | + | ++ | - | - | - | - | + | - | + | - | - | + | + |
| Hemimelia | + | ++ | - | + | + | - | + | - | + | - | + | + | + |
| luxation | + | ++ | - | + | + | - | + | - | + | - | + | + | + |
| Presacral vertebrae | d | d | - | - | - | d | d | i | i | - | - | d | d |
| Skull | - | - | + | + | - | - | - | - | - | - | + | - | - |
| Stomach | + | ++ | - | - | - | - | - | - | - | - | - | - | - |
| Spleen | ++ | ++ | - | - | - | - | - | - | - | - | - | + | - |
| Kidney | + | ++ | - | - | - | (+) | + | - | - | - | - | + | - |
| Reproductive system | - | + | - | - | - | - | - | - | - | - | + | - | + |
| Uro-rectal defects | - | - | - | - | - | - | - | - | - | - | - | + | - |
| Epidermal defects | - | + | + | + | - | - | - | - | - | - | + | - | + |
| Usual lethality | - | + | - | + | - | - | - | - | - | - | - | + | - |

luxoid genes. Secondly, and perhaps more interestingly, the luxoid mutations may disrupt a common developmental pathway, either at the molecular or cellular level. Isolation of these genes would therefore enable detailed molecular studies of the genetic interaction of the luxoid alleles. In addition the interaction of the luxoid mutations, alone or in combination, with other known limb genes can be investigated. Such studies may indeed establish a functional relationship between the luxoid genes which may be further integrated into the framework of events proposed for the genes already known to control limb development. As such, a combined investigation of genes which are related at the phenotypic level may provide a much more informative resource than the analysis of limb genes in isolation.

Recent studies illustrate the above point by demonstrating a relationship between preaxial polydactyly and the AP patterning properties of the ZPA. The formation of additional preaxial digits in polydactyly mutants was suggested to resemble the mirror image duplications which result from an ectopic ZPA in the anterior limb bud margin. Accordingly the mutants *Hx*, *Xt*, *Lx*, *Lst* ectopically express *Shh* and *Fgf4*, genes which mediate the action of the ZPA in the anterior mesenchyme (Chan *et al.*, 1995; Masuya *et al.*, 1995; Masuya *et al.*, 1997). Further Chan *et al*, (1995) show that the anterior mesenchyme of *Lst* limbs exhibits polarising activity when grafted into host chick limbs. These results demonstrate the multigenic control in the establishment of the AP axis in the mouse limb. Also it shows that indeed the luxoid genes can function in a common pathway of cellular events. Three of the luxoid mutants *Dh*, *Hx* and *Xt* are of particular relevance to this thesis, hence their phenotypes are described below.

### 1.3.2. *Dominant hemimelia*

*Dh*, which arose spontaneously, was discovered in 1954 by Carter but first described in detail by Searle, (1964). The external abnormalities of heterozygotes, which are confined to the preaxial side of the hind limbs show a high degree of penetrance (96% in Searles' study) but vary in severity. The phenotype in the digits

ranges from a slight thickening or triphalangy of the hallux to the addition of an extra digit, sometimes accompanied by syndactyly. Oligodactyly may occur with the loss of up to two digits. Tibial hemimelia, reduction and fragmentation of the femur and pubic element of the pelvis are seen and the number of ribs, sternebrae and presacral vertebrae may be reduced. It seems paradoxical that *Dh* can cause both polydactyly and oligodactyly and the causal mechanism for this remains unknown but Searle (1964) and later Rooze (1977) document alterations in limb bud morphology consistent with both cases. The first macroscopic distinction between *Dh*/+ and wild type limb buds can be seen at day 11.5. At day 12.5 heterozygous *Dh* limb buds are seen in which the preaxial end of the autopod, or footplate, is missing and there is a deep notch between the autopod and the distal part of the zeugopod. In addition the preaxial portion of the AER is lacking over a variable extent. It is likely that such limb buds correspond to limbs displaying absent preaxial prechondrogenic condensations, observed by Searle at day 14.5, which manifest as complete tibial hemimelia and oligodactyly in mature limbs. Conversely Rooze (1977) also document 11.5 day *Dh* heterozygous limb buds which appear to show an abnormal overgrowth of preaxial mesenchyme in the autopod, the zeugopod remaining normal. These cases are likely to correspond to limbs with a thickened or triphalangeous hallux but normal long bones.

Abnormalities are also seen in the viscera, the most prominent of which is absence of the spleen which is 100% penetrant. Only one other mouse mutant displays asplenia. Homozygosity for null alleles of *Hox11* leads to spleen agenesis but there is no evidence for additional abnormalities (Roberts *et al.*, 1994; Dear *et al.*, 1995). It is not clear whether in these mice the spleen fails to develop or whether spleen development is initiated but then fails due to apoptosis at day 13.5 (Roberts *et al.*, 1995; Dear *et al.*, 1995). Additional defects in *Dh* are small stomach, shorter than normal alimentary canal and a left kidney which is either flattened or occasionally hydrophic. Heterozygotes reach maturity and both males and females can breed.

Homozygotes present the same range of abnormalities but with increased severity, hence the mutation is considered semidominant. Heterozygotes rarely show overt urogenital abnormalities but in homozygotes defects include recto-vaginal fusion, recto-urethral fusion, absent or vestigial bladder, blind ending ureters, urethras or uterine horns. Few survive to weaning due to lethality caused by the visceral defects, but some achieve maturity and a female has even bred, showing that sterility is not a part of the *Dh* syndrome.

The earliest observed defects of *Dh* mice can be traced to defective splanchnic mesoderm in 9.5 day embryos (Green, 1967). At this stage the splanchnic mesoderm normally has an epithelial type of organisation which thickens from the anterior end to the primordium of the dorsal pancreas to form the anterior splanchnic mesodermal plate, ASMP. In both *Dh* heterozygotes and homozygotes the epithelial organisation of the ASMP is lost and the region of the splanchnic mesoderm posterior to the lung primordia is diminished or absent altogether. At the posterior end of the coelom, which seems not to elongate properly at this stage, mesoderm accumulates in an unorganised mass. The morphogenesis of the internal organs defective in *Dh* mice depend upon interaction with the splanchnic mesoderm, hence it seems likely that defective splanchnic mesoderm causes the abnormalities seen. A compounding factor is likely to be the shortened and abnormally shaped coelom causing aberrant spatial relationships of developing structures.

### 1.3.3 *Dh* is a gain of function mutation

*Dmh*, *Sdm* and *Swf* are three mouse lines which carry large chromosome 1 deletions (Cattanach *et al*, 1993). All show growth retardation, typical of such deletions, and *Dmh* has a domed skull and shortened face. However none manifest a *Dh* phenotype. The *Dh* locus must be deleted from *Dmh* and *Swf* because the *Dh* flanking markers, *En1* and the 3′ of *Gli2* are deleted. In *Sdm* however, although *Gli2* is deleted, *En1* and *Inhβb*, which lies within the *Dh* critical region (see chapter 3), are both present (David Hughes, manuscript in preparation). The *Dh* gene may or may

not therefore be deleted in *Sdm*. Because none of these lines manifest a *Dh* phenotype the mutation cannot be due to haploinsufficiency, rather it is a gain of function mutation.

*Dh* has been crossed to *Dmh*, *Sdm* and *Swf* to produce compound heterozygotes. All three genotypes present the same phenotype. A heterozygous *Dh* phenotype is seen with respect to the visceral defects, while the limb phenotypes are more similar to the *Dh* homozygote than the heterozygote (Jo Peters, pers. comm.). This data may suggest that the *Dh* limb phenotype is being enhanced due to the lack of another limb gene which is deleted from all three lines. *Gli2* would represent a good candidate as it is deleted from all three lines and a *Gli2* null allele is known to be recessive with effects on the skeletal patterning of the limb (Mo *et al.*, 1997; see chapter 4). Alternatively the enhanced *Dh* limb phenotype may be due to the deletion of the wild type *Dh* allele indicating that the *Dh* and wild type allele must interact. Were this to be the case then the visceral defects would be independent of the wild type allele and the phenotype would be dose dependent on the mutant allele. In addition it would suggest that the *Dh* gene must lie proximal to *Inhβb* because *Inhβb* is not deleted from *Sdm*.

### 1.3.4 *Hemimelic extra toes*

*Hx* is a semidominant mutation which arose spontaneously (Dickie, 1968). Heterozygotes are affected on all four feet, however the hind limbs are always more severely affected. The limb phenotype is fully penetrant but varies in expressivity. Typical expression includes hemimelia of the radius, tibia and talus, supernumerary metacarpals and metatarsals and preaxial polydactyly. The fibula and ulna are normal in size but usually bowed with an increase in bowing correlating to increased radial or tibial hemimelia. The humerus, femur and limb girdles are normal. No visceral abnormalities have been described. Heterozygotes are fertile but some males cannot breed due to crippled limbs. The homozygous condition is embryonic lethal at day 8 to 9 but no description of the phenotype has been reported.

Knudsen and Koc*hh*ar (1981) have examined limb-bud outgrowth in *Hx* heterozygotes. They observed an abnormal pattern of programmed cell death (PCD), or apoptosis, in distinct regions of the limb-bud. The preaxial apical ectodermal ridge, AER, did not undergo normal regression at stage 18-20 but remained in a thickened state. The preaxial mesoderm of the progress zone similarly displays a greatly reduced level of characteristic PCD. It is likely that the lack of PCD in the preaxial mesoderm in combination with a prolonged trophic influence from the AER leads to abnormal preaxial mesenchymal proliferation, producing the characteristic preaxial protrusion from which supernumerary digits form. The opaque patch is a zone of mesodermal cells which normally undergo PCD at stage 19-20 resulting in the separation of tibial and fibular precartillage rudiments. In *Hx* this zone is extended to the preaxial side resulting in a degeneration of the distal portion of the pre-cartilaginous tibia. This defect is thought to lead to tibial hemimelia.

Cells showing the characteristic morphology of normal PCD are seen in *Hx* mice, the mutation is therefore unlikely to affect the cell death mechanism itself. Rather the mutation results in a misplaced pattern of cell death. The mechanism which drives this aberrant PCD pattern is unknown but could be due to a defect in the normal pattern formation process itself or the mutation could affect a process, a secondary effect of which produces the abnormal pattern of cell death. Correspondingly, as mentioned above, Masuya *et al.*, (1995) report ectopic expression of *Shh* and *Fgf4* in anterior mesenchyme of day 12 limb buds indicating a duplicated ZPA.

A further limb mutation maps very close to *Hx* on chromosome 5. One recombinant has been found between *Hammertoe*, (*Hm*) and *Hx* in 3664 offspring suggesting that the genes are closely linked and possibly allelic (Sweet, 1982). *Hm* mice are characterised by feet in which the second phalanx of digits 2,3,4 and 5 is strongly flexed. The phenotype is caused by the failure of inter digital webbing to regress during development which prevents normal elongation of the digits. All four feet are affected with homozygotes showing a greater degree of webbing compared to

heterozygotes. Zakeri *et al.*, (1994) have shown a failure of normal PCD in the interdigital mesenchyme between digits two and five. PCD however progresses normally in the interdigital mesenchyme between digits one and two and in the anterior and posterior necrotic zones. As with *Hx* this implicates a defect controlling the pattern of PCD within the limb bud, rather than the PCD mechanism itself.

*Hx* and *Hm* are of direct clinical interest because three clinically distinct human polydactyly conditions, triphalangeal thumb (TPT), a complex bilateral polysyndactyly syndrome (CBP) and preaxial polydactyly type 2 (PPD-2) have been mapped to the human genome in a region of chromosome 7, 7q36, syntenic with the *Hx/Hm* region (Heutink *et al.*, 1994; Tsukurov *et al.*, 1994; Hing *et al.*, 1995). The three conditions segregate as autosomal dominant traits with a high degree of penetrance but variable expressivity. PPD-2 presents triphalangeal thumb with duplication of the thumb and big toe. TPT displays opposable and non-opposable triphalangeal thumbs in addition to mild cutaneous syndactyly and occasional postaxial polydactyly. CBP is characterised by triphalangeal thumbs and severe cutaneous syndactyly with pre- and postaxial polydactyly. Malformations of the feet were present in some affected cases, but usually were less severe than those observed in the hands. All affected individuals for all three syndromes present triphalangeal thumb but show variation of expressivity in the other affected areas.

Despite these phenotypic variations the three conditions may be allelic. Haplotype analysis demonstrates an identical haplotype between markers flanking TPT and CBP, suggesting a common founder. The PPD-2 haplotype for these markers is however different indicating an independent mutation, either in the same gene or a tightly linked gene (Hing *et al.*, 1995). Given the phenotypic similarities and the syntenic genetic localisation, the *Hx* and *Hm* genes represent strong candidates for these conditions.

### 1.3.5 *Extra toes*

*Xt* is also a semidominant mutation which arose spontaneously and has been described by Johnson (1967). In common with the luxoid group penetrance of the limb phenotype is almost complete but the expressivity is variable. Heterozygotes have a varying number of supernumerary digits on the pre-axial side of both fore and hind limbs. This ranges from a thickening or doubling of the claw of the hallux and pollux to one or two extra preaxial digits united by soft tissue syndactyly. A postaxial nubbin is occasionally seen on the fore- and hindlimbs. The long bones are unaffected. An interfrontal bone in the skull is seen in 90% of heterozygotes and hydrocephaly occurs in a small proportion. A white belly spot may also be present. Homozygotes have multiple abnormalities and die in utero or at birth. They can be first recognised at day 9 because of excessively large pharyngeal arches and a persistently open neural tube. They have paddle-shaped feet with up to eight or nine digits usually showing syndactyly. The radius is usually hemimelic while more severe hemimelia is seen in the tibia. Severe defects are seen in the axial and craniofacial skeletons with defects of the vertebrae, sternum and ribs and an enlarged maxillary region and open skull vault. The open neural tube results in abnormalities of the brain and sense organs (Franz, 1994; Franz and Besecke, 1991).

*Xt* is a mouse model for Greig cephalopolysyndactyly syndrome (GCPS). In 1926 Greig described an autosomal dominant disorder affecting limb and craniofacial development. Thumbs may be broad but are rarely duplicated. There is commonly a postaxial nubbin of tissue on the ulnar boarder of the fifth digit. Occasional syndactyly between fingers 3-5 is seen. In the feet the big toe is usually duplicated and there is syndactyly between the first two or three toes. Occasionally a postaxial digit is present. Craniofacial manifestations are variable and consist of a high forehead, hypertelorism, and scapho- or brachycephaly. The nose and nasal bridge is usually broad. Subsequently similar descriptions in several families have been reported (reviewed in McKusick, 1992). Winter and Huson (1988) compared this disorder to *Xt*, and concluded that the conditions are homologous. The phenotypic

similarities are pre-axial polydactyly of the hind limbs and a post-axial nubbin on the forelimbs. The broad forehead and prominent nose of GCPS patients are analogous to the large interfrontal bone seen in *Xt* heterozygotes. Hydrocephaly is also a common feature. *Xt* and GCPS are confirmed to be homologous conditions because both are caused by haploinsufficiency of *Gli3*. This is discussed in chapter 4, 4.1.2.

Recently a second human syndrome, Pallister-Hall syndrome (PHS), has also been shown to be caused by mutation in *GLI3* (Kang *et al.*, 1997). Both PHS and GCPS have polysyndactyly and abnormal craniofacial features but are clinically distinct, PHS presenting anomalies that are more severe than GCPS (Hall *et al.*, 1980; Clarren *et al.*, 1980). PHS presents hypothalamic hamartoma, postaxial polydactyly and imperforate anus. Some patients present additional defects including laryngeal cleft, abnormal lung lobation, renal agenesis, short 4th metacarpals, nail dysplasia congenital heart defect, multiple buccal frenula, hypoadrenalism, microphallus and intrauterine growth retardation. PHS does not cause hypertelorism or broadening of the nasal root or forehead which is seen in GCPS.

Frameshift mutations in *GLI3*, caused by single base pair deletions, were demonstrated in two PHS families. In both cases this would result in a protein truncated just 3′ of the zinc finger domain. The mode of action of such a truncated *GLI3* protein is unknown but due to the more severe phenotype seen in PHS compared to GCPS it is unlikely that the PHS alleles are simply causing haploinsufficiency, as is the proposed case for GCPS. The PHS alleles may be acting as antimorphs, in a dominant negative fashion. Alternatively the truncated protein may be acting as a neomorph by binding to additional promoter regions compared to the wild type *GLI3* thereby perturbing developmental pathways outside of the normal *GLI3* pathway.

### 1.4.1 The mouse as a model organism

The study of human biology relies on animal models to act as surrogates for experimentation which is either morally inconceivable or technically impossible in human subjects. Different biological problems are best studied in particular model systems but for many years the mouse has been well established as the model of choice for the study of mammalian genetics. This is due to three main reasons. Firstly, due to evolutionary conservation, the biology of the mouse is very similar to that of humans, although significant differences will inevitably be encountered (see Erickson, 1989). This means that meaningful deductions can be drawn about human biology from knowledge gained in the mouse. This is of the utmost importance in the efforts to model human disease in mice (Darling and Abbott, 1992). Secondly the mouse has become by far the most adept system technically for mammalian research. Intrinsically the mouse has a high fecundity, produces large litters and is relatively easy to maintain; factors which are crucial for the generation of sufficient progeny needed for genetic analysis. Also the generation of many hundreds of inbred strains and more recently wild derived strains enables controlled matings to be performed. However it is the enormous technical advances of recent years to enable high resolution genetic mapping and the ability to genetically manipulate the function of any gene which has truly singled out the mouse as the model of choice. Increasingly this is being illustrated by the modification of research paradigms, traditionally tailored for other creatures, for use with mice rather than attempting to reproduce the genetic technology of the mouse in other species. For example mammalian behavioural paradigms, traditionally the domain of the rat, are being redesigned to suit the ethological criteria of mice. This is allowing the effect of specific genetic manipulation, possible only in mice, to be studied at the behavioural level. Lastly, both the classical genetic methodologies and the modern molecular technology is economically and practically feasible for the whole community and resources can be easily disseminated to all laboratories.

### 1.4.2 Generating mutations in the mouse

A fundamental approach to the investigation of gene function is through the analysis of mutant phenotypes. Over 1200 mouse mutations, covering a wide range of phenotypes, have previously been catalogued (Green, 1989) but the mouse now offers unparalleled opportunities for the generation of novel mutations.

Many of the extant mutants have simply arisen spontaneously, the phenotype being noticed by an observant worker. All of the luxoid mutants arose in this way. Novel mutants have traditionally been created by the treatment of mouse germ cells with any one of a number of potent mutagens such as ionising radiation (Green and Roderick, 1966) or chemical agents such as ethylnitrosourea (ENU) or chlorambucil (CHL) (Rinchik, 1991). These treatments result in genetic lesions which occur randomly throughout the genome. Typically it is then necessary to screen large numbers of progeny to detect phenotypes of interest. Intrinsically these methods will preferentially detect semidominant mutations, especially those which display an overt, visible phenotype because no further breeding regime is required. However to uncover recessive mutations or embryonic lethals careful breeding protocols are required (Brown and Peters, 1996: Rinchik *et al.*, 1990). This disadvantage has resulted in the mouse mutant resource being poorly represented for recessive mutations, while embryonic lethals are virtually non-existent. Indeed a gulf exists between the present mouse mutant resource and the full range of phenotypes which are possible. Only 1-2% of the total number of mouse genes have yielded mutant loci and many phenotypes which will model the full range of human genetic disease remain to be established. Brown and Peters (1996) have termed this deficiency the 'phenotype gap' and describe efforts to narrow this gap through saturation ENU mutagenesis. Extensive breeding protocols using the stock of mouse lines that carry large deletions and an integrated approach to screening for diverse phenotypes will be required to increase the efficacy of such a programme. The potential of this approach is illustrated by the recovery of recessive mutations in the albino-deletion region of

mouse chromosome seven (Rinchik *et al.*, 1990) and the recent identification of the circadian rhythm gene *Clock* (Vitaterna *et al.*, 1994).

Novel mutations may be created by one of several molecular genetic techniques. Again in a random fashion DNA can act as an insertional mutagen during the production of transgenic mice (Jaenisch, 1988; Meissler, 1992). Because the transgene itself tags the insertion site a valuable handle is provided for the molecular cloning of the disrupted gene. Transgene insertion however frequently induces complex intra-chromosomal rearrangements and deletions which can greatly hinder the molecular analysis.

In a more directed approach transgene vectors have been developed which specifically target genes, albeit randomly (Joyner *et al.*, 1992). These strategies are typically based on the electroporation into embryonic stem, ES, cells of β-galactosidase reporter constructs which lack a functional promoter (Gossler *et al.*, 1989; Friedrich and Soriano, 1991). Only when the transgene integrates into a gene undergoing transcriptional activity (concomitantly inactivating and tagging the gene) can β-galactosidase be expressed. Mutant ES cells are then used to generate chimeric embryos which can be screened for genes of interest by analysis of the β-galactosidase expression pattern. Ultimately heterozygous and homozygous animals are generated to determine the mutant phenotype, while molecular analysis of the integration site will identify the mutated gene.

All the above strategies are phenotype driven, the production of an interesting phenotype directs investigation towards the mutated gene. Conversely a cloned gene may be specifically mutated via the technology of gene targeting (Capecchi, 1989; Joyner, 1993). This immensely powerful approach relies on the combination of several technologies including the ability to propagate embryonic stem (ES) cells *in vitro* such that they retain totipotency, the capacity to induce and select for homologous recombination events which generate the desired mutation within ES cells and the ability to generate chimeras from targeted ES cell clones. To date the most common application of gene targeting has been to generate loss of

function or 'knockout' mutations. However protocols also exist to generate subtle lesions, such as point mutations or site specific changes in regulatory elements (Joyner, 1993). This will enable the generation of gain of function mutations, increasing the depth of the mouse mutant resource, and enable specific aspects of gene product function to be dissected. Importantly, the ability to create mutations to order will enable genetic lesions responsible for human disease to be reproduced in mouse models which can be extensively analysed.

A major drawback of conventional gene targeting technology is that the engineered mutation is under no spatial or temporal control; the mutation will be present in every cell of the animal, throughout the animals life. This may complicate or often preclude meaningful analysis of particular aspects of gene function. For example many genes act at multiple times and in multiple areas during development and in the adult. Roles during later developmental events, such as organogenesis, or physiological roles in the adult may not therefore be discernible if the animal has been unable to execute its normal developmental program. Recent developments of gene targeting technology however promise the exciting possibility of conditional gene targeting whereby mutation can be induced in a spatially and temporally defined manner. The phage P1 derived *Cre-loxP* recombination system combined with the ES cell gene targeting technology potentially allows the production of a specific mutation restricted to one particular cell type (Gu *et al.*, 1994). Further developments will enable the *Cre-loxP* system to be placed under inducible control enabling the engineered mutation to be generated at any desired time (e.g. Kuhn *et al.*, 1994; Shockett and Schatz, 1996; No *et al.*, 1996).

It is now also possible to engineer specific mutations via the production of transgenic mice using large genomic fragments, in the form of yeast or bacterial artificial chromosomes of up to 670kb (reviewed by Peterson *et al.*, 1997). Transgenes can now be introduced in the context of their native genomic locus, with the inclusion of remote regulatory elements enabling faithful recapitulation of endogenous expression. This was only rarely possible using conventional techniques

due mainly to the size constraints (40-50kb) of phage and cosmid vectors. With the ability to engineer site directed mutation into YAC clones using the homologous recombination capacity of yeast, specific mutations can be readily produced in YAC clones which can then be introduced into the mouse germline. As the transgene is likely to be expressed correctly *in vivo* the phenotypic consequences due to the mutation can be more readily assessed.

The techniques for the manipulation of the mouse germ-line are now widely employed and the number of chemically/ionising radiation induced, transgenic, gene trap and knockout mice is increasing exponentially. While much can be learnt from individual mutants alone the long term value of this resource will derive from interbreeding to produce compound mutant lines. This will enable the functional analysis of gene interaction, the ordering of genes in genetic pathways and the functional dissection of gene families displaying apparent gene redundancy.

### 1.4.3 Cloning mutant genes in the mouse

Traditional techniques for cloning mutations in the mouse and in human have relied on a degree of biological information concerning the gene product involved. These strategies have involved protein purification followed by peptide sequence determination or utilised the intrinsic functional properties of a gene to facilitate cloning. However for the vast majority of single gene mutations no biological information concerning the basic genetic defect exists. Indeed, as recently as 15 years ago it was unclear if the genes for such disorders could be identified in the absence of direct information about the biological defect. However, in the mouse, the advances in molecular genetics of the past decade now make it feasible to clone any mutation. This new technology is termed 'positional cloning' to reflect that any gene can be isolated based on knowledge of its chromosomal location, in the absence of any biological information (Collins, 1992). Positional cloning has resulted from the concomitant advances in a number of fields including the generation of a high resolution, marker dense genetic map and the ability to clone the large physical

regions to which a gene can be genetically defined. In addition many approaches are now available for the isolation of genes within such tracts of genomic DNA. Of the >900 mouse mutations identified, fewer than 100 have been cloned (Dietrich *et al.*, 1995), however all these genes are now potentially accessible via positional cloning.

The initial step in a positional cloning project is to locate the mutation to a chromosomal position relative to previously placed markers. This is achieved by the generation of a mapping panel in which the mutant phenotype is segregating, usually via either the backcross or intercross protocol (see Silver, 1995). This task has been greatly aided by two developments of methodology. Firstly mapping was traditionally performed on intraspecific crosses or recombinant inbred strains generated using laboratory strains of *Mus musculus* (Taylor, 1989). However the use of these crosses is limited by the low levels of interstrain polymorphism which can severely curtail the number of markers which can be followed. This problem was overcome by Bonhomme and colleagues who devised the use of distinct mouse species for the generation of interspecific crosses (Bonhomme *et al.*, 1979; Avner *et al.*, 1988). The evolutionary distance between *Mus musculus* and the related species *Mus spretus* is sufficient to detect high rates of polymorphism but still allows the production of fertile $F_1$ hybrids. Potential problems in crossing *Mus spretus* with *Mus musculus* strains however include poor breeding with infrequent, small litters, sterile $F_1$ males and the existence of small inversions that will act to eliminate recombination and therefore distort the true genetic map (Hammer *et al.*, 1989). To overcome these disadvantages the *Mus musculus* subspecies, *M. m. castaneus*, is commonly used to generate crosses in combination with *M. m. domesticus* or *M. m. musculus* derived strains.

The second major advance has been the discovery and exploitation of a highly abundant and polymorphic class of loci known as microsatellites or simple sequence length polymorphisms (SSLPs) (Weber and May, 1989). These markers are anonymous and consist predominantly of di-, tri-, or tetranucleotide repeats. The most common repeat in the mouse genome is the dimer (CA)n:(GT)n, estimated at an

occurrence of one every 18kb (where n is greater than six) (Stallings *et al.*, 1991). The polymorphic nature of these sequences is due to mispairing or slippage during replication or recombination resulting in an expansion or contraction in the number of repeats. The rate of polymorphism varies with the size of the repeat; larger, uninterrupted repeats showing greater polymorphism than smaller repeats. In an analysis of over 300 CA repeat loci (where n > 15) an average polymorphism rate of approximately 50% was observed in pairwise comparisons among nine classical *M. musculus* inbred strains while the enhanced levels of polymorphism inherent between *Mus* species or subspecies was also demonstrated. The rate of polymorphism between B6 and *M. m. castaneus* was 77% and between B6 and *M. Spretus* was 90% (Love *et al.*, 1990; Dietrich *et al.*, 1992). The power of microsatellite markers lies not only in their abundance and high rate of polymorphism but also in the ease with which they can be typed (by PCR), uncovered and disseminated.

SSLP loci have been exploited to generate a dense genetic map of the mouse. Over 6600 SSLPs have been isolated and mapped throughout the genome (Dietrich *et al.*, 1994). This represents, on average, a resolution of one marker every 0.25cM or 500kb. Further a subset of these markers have been designated as anchor loci which are evenly spaced, at between 10 to 20cM intervals. These markers enable the integration of the different crosses such as the Whitehead Intercross (Dietrich *et al.*, 1992), the NCI/Frederick Interspecific Backcross (Copeland *et al.*, 1993) and the European Collaborative Interspecific Backcross (EUCIB) (EUCIB Group, 1994) which have, and are continuing to be used, to generate a genome wide linkage map. The plethora of information assimilated within these maps is now available via the Internet at http://www-genome.wi.mit.edu for the Whitehead and NCI/Frederick maps and at http://www.hgmp.mrc.ac.uk/mbx/mbxhomepage.html for the EUCIB map. (The use of EUCIB is discussed further in chapter 6).

With the tremendous density of markers now placed on the mouse genome, over 14,000, (Dietrich *et al.*, 1995) it becomes possible to define a mutation to a minimal genetic region. This then allows previously mapped genes which also map to

the region of the mutation to be assessed for concordance with the mutation. If a gene fails to recombine with the mutation then it can be considered a candidate gene for the mutation. Functional knowledge about the gene in relation to the aetiology of the mutation will either enhance or diminish the strength of candidacy. The majority of mouse mutations cloned to date have been cloned by this 'candidate' approach including *brachypodism* (*bp*) in which *Gdf5*, a member of the BMP family of secreted signalling molecules, is mutated (Storm *et al.*, 1994) and *Splotch* (*Sp*) caused by mutation in the transcription factor *Pax3* (Epstein *et al.*, 1991). Given the rate at which new genes are being placed on the mouse and human genome maps, particularly expressed sequence tags (ESTs), doubtless many more will be cloned by this approach (Collins, 1995).

This rationale can be extended to comparative genomics. Approximately 3500 genes have been mapped in the mouse, of which approximately 40% have also been mapped in human. These genes define about 110 blocks of synteny conservation which are estimated to cover between 80% and 90% of the mouse autosomal genome (Dietrich *et al.*, 1995). It is therefore possible to predict the existence and location of a gene in either species, based on the presence of an orthologous gene in the other. Further genes identified from mouse mutations can be considered candidates for human disorders presenting a comparable phenotype and viceversa. For example, identification of a type VII myosin gene causing the mouse mutation *shaker-1* (Gibson *et al.*, 1995) led rapidly to the identification of a human homologue which was responsible for Usher syndrome type 1B (Weil *et al.*, 1995).

Definition of a regional genetic map encompassing a mutant allele also serves as the foundation for the generation of a physical map. This strategy becomes essential if the mutant gene cannot be determined by the candidate gene approach. Typically the genetic interval in which a mutation must lie can be resolved to 1cM or less, representing, on average, a physical distance of less than 2Mb. Crucially it is possible to clone these large physical distances due to the development of yeast artificial chromosome (YAC) vectors which can carry up to 2Mb of insert DNA. In

addition other large insert vectors have been developed including bacterial artificial chromosomes (BACs) and P1-derived artificial chromosomes (PACs) which have a cloning capacity of up to 300kb (reviewed by Monaco and Larin, 1994). YAC vectors contain the basic functional units of yeast chromosomes: a centromere, two telomeres and autonomous replication sequences (Burke *et al.*, 1987). These elements enable insert DNA to be propagated within yeast as an artificial chromosome. Libraries of the mouse genome are available to the community (Chartier *et al.*, 1992; Larin *et al.*, 1991; Kusumi *et al.*, 1993) and can be easily screened to isolate YACs containing markers which flank a mutation. With luck these clones may already overlap and if not the series of overlapping YACs, or contig, may be completed by one or more steps of chromosomal walking using markers derived from previously isolated YACs (see chapter 3, 3.2.3-3.2.4). Frequently YAC derived markers can also be mapped relative to the mutation enabling refinement of the maximal physical region in which a mutation can lie. Several major problems can however be encountered when working with YACs. Firstly a high percentage of clones are chimeric, containing two separate segments of non-contiguous DNA. Secondly some clones are unstable and undergo internal deletions during propagation and thirdly YAC DNA cannot be easily purified from the yeast host chromosomal DNA. The BAC and PAC cloning systems, being bacterial-based, may alleviate these short comings.

Refinement of the maximal physical region in which a mutation can lie can be achieved by the recent developments in the ability to express large insert genomic clones in transgenic mice (Peterson *et al.*, 1997). In this way it is possible to assay for complementation of a mutant phenotype by introducing large insert clones containing candidate regions into the germline of the mutant strain. Antoch *et al.*, (1997) have used this approach to facilitate cloning of the circadian rhythm gene *Clock*. The mutant *Clock* allele is antimorphic, the phenotype was therefore rescued by overexpression of the wild type gene. Neomorphic and hypermorphic mutations however will not be rescued by complementation with the wild type gene, however

this approach could still be used for such alleles by reproducing the mutant phenotype using genomic clones derived from the mutant DNA.

As a long term strategy of the Mouse Genome Project efforts are underway to assemble genome-wide physical maps by isolating YACs corresponding to each of the SSLPs in the genetic map. With SSLPs placed, on average, every 450kb and a YAC library with an average insert size of 800kb, the intervals between markers will typically be closed by YACs. Such a resource will considerably facilitate positional cloning as YAC contigs for a specific regions will be, at least partially, obtained simply by screening the appropriate database (http://www-genome.wi.mit.edu.).

The rate limiting step in positional cloning projects is frequently the identification of candidate genes within the large physical region of the YAC contig. Numerous strategies are available including exon trapping, cDNA selection, CpG island isolation, and direct sequencing (reviewed by Parrish and Nelson, 1993). Each method presents advantages and disadvantages and typically a combination of methods are employed.

### 1.5 Goals

The ultimate aim of the work presented in this thesis was to isolate the genes responsible for the mutations *Dh* and *Hx*. The *Dh* mutation had previously been located to a genetic interval of 1.2±0.5 cM by backcross analysis. Cloning of the *Dh* gene was therefore to be approached via the positional cloning strategy of isolating the physical region represented by this genetic interval. As presented in chapter 3, the initial objective was the generation and characterisation of a complete YAC contig encompassing the two closest flanking markers to *Dh*. From this resource it was hoped to isolate candidate genes for *Dh* which would be tested by mutational analysis. Concomitantly candidate genes arising in the literature were to be assessed for a possible involvement in *Dh*, as described in chapters 3 and 4.

A genetic resource providing tight linkage of *Hx* to flanking markers was not available. However *Dh* and *Hx* are postulated to reside in paralogous linkage groups

suggesting a common ancestry for the two genes. The candidacy of *Gli2* for *Dh* raised the possibility that a novel *Gli* gene was involved in *Hx*. The work presented in chapter 5 details the aim of cloning additional *Gli* family genes which would then be assessed for involvement in *Hx*.

Finally chapter 6 describes the chromosomal location, developmental expression pattern and biological role of *Pmsc2* which suggested candidacy for *Hx*. The aim of this work was therefore to assess this candidacy by high resolution genetic mapping of *Pmsc2*.

# CHAPTER 2

# MATERIALS AND METHODS

All chemicals were supplied by BDH unless otherwise stated. All solutions and methods were taken from or adapted with minor modifications from Sambrook *et al.* (1989) unless otherwise stated.

## 2.1 Mouse DNA samples

BXD recombinant inbred strain DNA and BSS backcross DNA samples were obtained from The Jackson Laboratory, Bar Harbor, ME, USA. EUCIB DNA samples were obtained from the HGMP Resource Center, Hinxton Hall, Cambs.

DNA was prepared from animals carrying the informative Dh chromosome generated by the *Dh* backcross (Higgins *et al.*, 1991) from frozen tissue using method 2.3.5. The animals, carrying informative recombinations, used to prepare DNA had been maintained in the following way: 441 was the original backcross animal, the recombinant Dh chromosome against the LIII chromosome; The recombinant Dh chromosome of animals 194, 386, 473 and 481 had been bred to homozygosity and the recombinant Dh chromosome of animals 268, 299 and 345 was maintained on a C57BL/6 background.

## 2.2 Bacterial Cell Culture

### 2.2.1 Media and solutions

Luria Bertani, LB medium

In 1 liter water dissolve 10g tryptone (Difco); 5g yeast extract (Difco); 10g NaCl. Sterilise by autoclaving.

LB-agar

In 1 liter water dissolve 10g tryptone (Difco), 5g yeast extract (Difco), 10g NaCl, 15g agar (Oxoid Ltd). Sterilise by autoclaving. When needed ampicillin, X-Gal and IPTG were added to molten media (below 48°C) prior to pouring of plates.

Ampicillin (Sigma)

Make stock at 50mg/ml in $dH_20$, filter sterilise. Dispense into 1ml aliquots and store in the dark at -20°C. Use at final concentration of 100μg/ml.

X-Gal (Sigma)

Make stock at 20mg/ml in dimethylformamide (Sigma). Store protected from light at -20°C. Use at a final concentration of 40μg/ml.

IPTG (Sigma)

Make stock at 100mM. Store protected from light at -20°C. Use at a final concentration of 0.5mM.

Freezing medium

50% glycerol in LB medium. Sterilise by autoclaving.

### 2.2.2 Growing bacterial cells on agar plates

The desired volume (up to ~200μl) of bacterial culture was pipetted onto the surface of an LB-agar plate (containing ampicillin, X-Gal and IPTG when desired) and spread with a sterile bent glass rod until the liquid had thoroughly soaked into the agar. (If more cells than are contained in 200μl were required, the cells were concentrated by a 2 second pulse of centrifugation in a microfuge followed by resuspension of the pellet in a smaller volume). The plates were then inverted and

incubated for 12-16 hours at 37°C. Cells were viable from plates stored at 4°C for several weeks.

### 2.2.3 Long term storage of bacteria

The selected clone was freshly struck out on an LB-agar plus ampicillin plate and grown overnight at 37°C. A single colony was then picked to inoculate 2ml of LB plus ampicillin medium. The culture was grown overnight to approximately mid-log phase at 37°C, shaking at 270rpm. 1.4ml of cell culture was then mixed with 0.6ml freezing medium, to give a final glycerol concentration of 15% and dispensed into sterile 1.8ml freezing vials (Treff). These were then frozen and stored at -70°C. Bacteria prepared in this way can be stored indefinitely.

### 2.2.4 Preparation of competent cells

Using a sterile loop, XL1-blue bacterial cells (Stratagene) from a frozen stock were streaked out onto an LB-agar plate and grown overnight at 37°C. 10ml of LB medium was then inoculated with a single colony and grown overnight in a shaking incubator, 250rpm, at 37°C. This culture was then used to inoculate 1000ml of LB which was grown to mid-log phase at 37°C, shaking at 250rpm; until an $O.D._{600nm}$ reading of between 0.5 and 1.0 was reached. The flask containing the culture was then chilled on ice for 15-30mins and the cells then harvested by centrifugation in ice cold 50ml Falcon tubes for 15mins at 4000rpm at 4°C (in a Sorval T6000). The supernatant was discarded and the pellets resuspended in a total of 1000ml sterile water at 4°C. Cells were then collected by centrifugation as above and resuspended in a total of 500ml sterile water at 4°C and re-centrifuged. Pellets were then resuspended in 20ml of 10% glycerol and centrifuged again. Cell pellets were finally resuspended in 3ml of 10% glycerol and aliquoted into 50µl volumes in 1.5ml eppendorf tubes and snap frozen in a dry ice plus ethanol bath. The aliquots were stored at -70°C. Competence of the cells was tested by transformation with

pBluescript vector alone, as measured by colony forming units, c.f.u., obtained per μg of transformed DNA. Values of $10^7$ to $5x10^8$ c.f.u./μg pBluescript were routinely achieved.

### 2.2.5 Transformation of competent cells

A 50μl aliquot of competent cells was thawed on ice, the transforming DNA (a maximum of 50ng DNA in a maximal volume of 2μl) added, gently mixed and incubated on ice for 1min. This was then transferred to an ice-cold cuvette (Flowgen) and subjected to a pulse of 2.5kV, at 25μF, 200Ω, in a BioRad Gene Pulser. The cells were swiftly mixed with 1ml of LB, supplemented with 10mM $MgCl_2$, transferred to an eppendorf tube and incubated at 37°C for 1hour. Aliquots of several different volumes were then spread onto LB-agar plus ampicillin plates and incubated overnight at 37°C.

### 2.2.6 Identification of colonies containing recombinant plasmids

Three different approaches were employed for the identification of recombinant plasmids, the method used depending on the nature of the clone to be identified.

(i) Selection by α-complementation.

Many plasmid cloning vectors in common use, including pBluescript (Stratagene) and pCRII (Invitrogen) are amenable to screening by α-complementation. Clones carrying DNA inserts, when plated on LB-agar plates containing X-gal and IPTG, are white, while non-recombinant clones are blue. White colonies are therefore picked, and grown overnight in 2ml LB + ampicillin cultures. DNA is then prepared (see section 2.3.1) and the desired clones identified by restriction analysis.

(ii) Selection by PCR analysis.

Rapid identification of recombinant clones can be achieved by PCR amplification of cloned DNA from individual colonies. Cloned DNA fragments can be amplified using primers complementary to vector sequences, such as the T3 and T7 primers which flank the multiple cloning site of pBluescript. Clones containing the expected sized product are identified and the amplified product is amenable to further analysis. Alternatively primers complementary to the DNA fragment being cloned can be used to identify recombinant clones. The protocol to PCR amplify directly from bacterial colonies is detailed in section 2.9.1.

(iii) Screening by hybridisation.

The method of Buluwela et al., (1989), an adaptation of Grunstein and Hogness (1975) is the method of choice for screening large numbers of recombinant clones. Bacterial colonies are lifted from LB-agar plates onto nylon membranes (Amersham) by placing dry filters directly onto plates. For orientation, a syringe needle charged with Indian ink is stabbed through each filter and plate in four asymmetric locations. Filters are then peeled off and placed, colony face up, onto Whatman No1 paper soaked in 2 x SSC / 5 x SDS and left for 2 minutes. Agar plates were then incubated at 37°C until colonies had regrown. The Whatman paper with filters is then transferred to a microwave oven with a rotating turntable and heated for 3 minutes at full power (650 Watts). Filters are then rinsed in 2 x SSC and subjected to standard prehybridisation for 30 minutes, hybridisation for 4 hours and normal washes. Following autoradiography, typically for one hour, plates are aligned and colonies corresponding to positive signals picked for further analysis.

## 2.3 DNA isolation and purification

### 2.3.1 Small scale preparation of plasmid DNA

For the isolation of up to 10μg plasmid DNA, bacterial cultures were grown overnight at 37°C, shaking at 250rpm, in 2ml LB plus ampicillin. Cells were

transferred to a 1.5ml microfuge tube and harvested by centrifugation. Plasmid DNA was then purified using the Wizard™ Miniprep system (Promega), according to manufacturers instructions.

### 2.3.2 Large scale preparation of plasmid DNA

For the isolation of up to 1mg plasmid DNA the selected clone was grown overnight at 37°C, with shaking at 250 rpm, in 2ml LB medium plus ampicillin. This culture was then used to inoculate 400ml LB plus ampicillin and incubated overnight at 37°C, with shaking at 250rpm. Cells were harvested by centrifugation at 4000rpm for 15 minutes at room temperature. Plasmid DNA was then purified using the Wizard™ Maxiprep system (Promega), according to manufacturers instructions.

### 2.3.3 Purification of DNA from agarose and aqueous solutions

DNA fragments to be recovered from agarose were viewed under a long wave (305nm) U.V. transilluminator and excised in a minimum of gel using a sterile scalpel blade. DNA was then isolated using Geneclean (Bio 101 Inc.) according to manufacturers instructions. The recommended Geneclean procedure was also followed for the purification of DNA in solution.

Extraction with phenol:chloroform followed by ethanol precipitation was also used to purify DNA from aqueous solutions. The salt concentration is initially adjusted to 0.2M NaCl by the addition of an appropriate volume of 5MNaCl. An equal volume of phenol:chloroform was then added to the DNA sample in a polypropylene tube and shaken vigorously. The sample was then centrifuged to achieve separation of the organic and aqueous phases. The aqueous phase was then transferred to a clean tube taking care not to take any debris at the interface. This procedure was repeated until the interface appears completely clear of debris. The aqueous phase is then transferred to a clean tube and 2 volumes of ethanol added. After 10 minutes on ice the DNA was recovered be centrifugation at 12000g for 10 minutes at 4°C (if the DNA was at a high concentration it could be recovered without

centrifugation by spooling with a sealed glass pipette). The DNA pellet was then rinsed in 70% ethanol, recovered by centrifugation and air dried until all traces of ethanol had gone. The DNA was then resuspended in an appropriate volume of the desired buffer.

### 2.3.4 Purification of DNA from polyacrylamide

The DNA band of interest was located by autoradiography and the band cut out using a sterile scalpel blade. In a 1.5ml microfuge tube the gel slice was crushed against the wall of the tube using a disposable pipette tip. Two volumes of elution buffer (0.5M ammonium acetate, 10mM magnesium acetate, 1mM EDTA (pH 8.0), 0.1% SDS) were then added. The sample was then incubated at 37°C on a rotary wheel for four hours (DNA fragments >500bp require a 12 hour incubation). The sample was then centrifuged at 12000rpm at 4°C for 1 minute and the supernatant transferred to a fresh tube. A 0.5 volume of elution buffer was added to the polyacrylamide, vortexed briefly, recentrifuged and the supernatant combined with that previously recovered. Any remaining fragments of polyacrylamide were removed by centrifugation at 12000rpm, 4°C for 10 minutes. DNA was then recovered by ethanol precipitation as described in section 2.2.3.

### 2.3.5 DNA extraction from internal mouse organs

Approximately 0.5g of tissue (from spleen, kidney or liver) was homogenised in 2.5ml TNES (50mM Tris.HCl (pH 7.5); 0.1M EDTA (pH8.0); 0.4M NaCl; 0.5% SDS). For DNA preparation from frozen tissue approximately 0.5g tissue was ground to a fine powder using a mortar and pestle prechilled to -70$^0$C. The powder was then taken up in 2.5ml of TNES. Proteinase K was then added to 100μg/ml and incubated overnight at 55°C. The homogenate was then split into 3 microfuge tubes, 0.5ml 2.6M NaCl added and the samples centrifuged at 12000rpm for 10 minutes. The supernatants were then pooled and two volumes of ethanol added. The DNA

was spooled out using a sealed glass pipette and rinsed in 70% ethanol. The DNA was left to air dry and then resuspended in 500µl TE (10mM Tris.HCl (pH8.0); 1mM EDTA).

### 2.3.6 Preparation of mouse genomic DNA in agarose plugs

Freshly collected tissue (spleen or liver) was cut into small pieces and homogenised in ice cold PBS (0.14M NaCl, 3mM KCl, 10mM $Na_2HPO_4$, 2mM $KH_2PO_4$, pH7.4). Fragments of connective tissue were removed by filtration through two layers of cheesecloth. Cells were then washed three times in ice cold PBS and resuspended at a concentration of 5 x $10^7$cell/ml ice cold L buffer (0.1M EDTA (pH 8.0), 0.01M Tris.HCl (pH 7.6), 0.02M NaCl). This cell concentration results in 100µl plugs containing approximately 10µg DNA.

The cell suspension was then warmed to $42^0$C and an equal volume of 1% low melting point agarose (Ultrapure LMP agarose, Gibco BRL) in L buffer, also at $42^0$C, added. After gentle but thorough mixing the cell suspension was dispensed, in 100µl aliquots, into plug molds (plug molds were wells of a 96 well plate, the plastic bottoms of which had been removed. Wells were thoroughly washed and sealed with tape). The plugs were placed on ice until set and then expelled from the mold into 50 volumes of L buffer containing 1mg/ml proteinase K and 1% sarkosyl. Plugs were then incubated at $50^0$C for 24 hours, the digestion mixture replaced with fresh digestion mixture and incubated at $50^0$C for a further 24 hours.

Plugs were washed three times in a large excess of TE, (pH 7.6) (10mM Tris.HCl (pH 7.6), 1mM EDTA) and then incubated for 1 hour at $50^0$C in 50 volumes of TE containing 40µg/ml PMSF. The plugs were then placed in 50 volumes of fresh TE/PMSF solution and incubated for a further hour at $50^0$C. Plugs were stored at $4^0$C in 0.5M EDTA (pH 8.0).

## 2.4 Enzymatic manipulation of DNA

### 2.4.1 Restriction enzyme digestion of DNA in solution

Digestion of plasmid DNA and purified DNA fragments were carried out with restriction enzymes (Boehringer Mannheim, NEB and Gibco BRL) according to manufacturers instructions.

Digestion of PCR product, which did not require a gel purification step was also carried out following manufacturers instructions. Prior DNA purification was not performed. When using enzymes requiring a low salt buffer the volume of reaction buffer included in the reaction was adjusted accordingly.

Digestions of genomic DNA in solution were carried out in a total volume of 50μl containing 10μg of DNA, 1 x the appropriate restriction buffer, 100μg/ml BSA and 10 units of enzyme. A further 10 units of enzyme were added after 4 hours. Reactions were incubated overnight at the appropriate temperature.

### 2.4.2 Restriction enzyme digestion of DNA in agarose plugs

Plugs, stored in 0.5M EDTA (pH 8.0) were washed by gentle agitation in 100 volumes of TE at room temperature for 10 mins. This was repeated six times. Plugs were then soaked in 30 plug volumes of 1x restriction enzyme buffer, supplemented with 200μg/ml BSA for 1 hour at room temperature, with frequent inverting.

For each digestion, a single plug was placed in an eppendorf tube in a total volume of 100μl containing 1x restriction buffer, 200μg/ml BSA and 20 units of restriction enzyme. Digestion was carried out overnight at the appropriate temperature. Following digestion the reaction solution was removed and the plug rinsed in 1ml ice-cold TE. Plugs were then kept on ice in 1ml TE prior to gel loading.

### 2.4.3 Dephosphorylation of 5′ termini

Dephosphorylation of the 5′ termini of linearised cloning vector, to prevent recircularisation of vector molecules only during ligation reactions, was achieved using Calf intestinal phosphatase, CIP, (Boehringer Mannheim) according to manufacturers instructions.

### 2.4.4 Phosphorylation of 5′ termini

Polynucleotide kinase, PNK, (Boehringer Mannheim) was used, according to manufacturers instructions, for the addition of the γ phosphate group of ATP to the 5′ OH-termini of DNA. It is necessary to kinase oligonucleotides before they can be ligated to restriction enzyme digested DNA.

### 2.4.5 Generation of blunt end DNA

The generation of blunt ended DNA fragments, to enable cloning or the addition of blunt ended linkers was achieved using T4 DNA polymerase (Boehringer Mannheim) according to manufacturers instructions.

### 2.4.6 Ligation of DNA

On ice insert DNA and vector DNA were mixed, typically in a molar ratio of 3:1. DNA concentration was estimated by estimation of band intensity compared to known quantities of molecular weight marker following agarose gel electrophoresis. Typically 10-100ng of vector DNA was used. The reaction was carried out according to manufacturers instructions in 10-20μl using T4 DNA ligase (Boehringer Mannheim). For ligation of cohesive termini 0.1 unit/10μl reaction was used while for blunt ended fragments 5 units/10μl reaction was used. Reactions were performed at room temperature for 1 hour or overnight (16 hours) at 16°C. If required the enzyme could be inactivated by heating to 65°C for 15 mins.

Cloning of PCR products was performed by direct ligation into the plasmid pCRII (Invitrogen), according to manufacturers instructions.

### 2.4.7 Shotgun cloning of lambda clone DNA into pBluescript

2μg of purified lambda DNA was digested to completion with the appropriate enzyme. This was checked, by agarose gel electrophoresis of a small aliquot, for the production of a smear of DNA fragments of the expected size; for example *Sau*3AI is expected to yield fragments of 300-350bp. DNA was then purified using the Geneclean procedure and ligated into appropriately linearised pBluescript at a molar ratio of 3:1. Typically 30ng digested lambda DNA is ligated into 100ng pBluescript. Following transformation cells are plated at low density (transformed cells recovered in 1ml LB medium are spread onto twelve 90mm LB/ampicillin plates) to facilitate screening by hybridisation.

## 2.5 Electrophoresis

### 2.5.1 Electrophoresis solutions and size markers

20x TBE

1M Tris.HCl (pH 8.0); 20mM EDTA, 1M boric acid (pH 8.3).

10x DNA loading buffer

20% Ficoll (Pharmacia), 100mM EDTA, Orange G (Sigma).

6% denaturing polyacrylamide

Pre-prepared acrylamide/bis-acrylamide (19:1w/w) sequencing solution; 6% w/v acrylamide, 7M urea and 1xTBE (Severn Biotech Ltd.) was used. 60mls of this solution was prepared to pour each 40cm x 30cm x 0.4mm gel. 75μl of TEMED (Sigma) and 0.5ml of freshly made 10% ammonium persulphate were added to the acrylamide before pouring the gel.

6% non-denaturing polyacrylamide

For 50ml of solution 7.5ml 40% 29:1 acrylamide:bis-acrylamide solution (EasiGel, Scotlab), 10ml 5 x TBE and 32.5ml water were combined. Immediately before casting the gel 60μl TEMED and 0.4ml fresh 10% ammonium persulphate were added.

DNA size markers

For agarose and 6% non-denaturing polyacrylamide gels 250ng of the appropriate size marker was used:

a) λ DNA digested with *Hind*III (Boehringer Mannheim)

b) ϕX174 digested with *Hae*III (Boehringer Mannheim)

c) 1kb ladder (Gibco BRL)

Lambda DNA concatemer size markers (NEB) were used for pulsed field gel electrophoresis.

### 2.5.2 Agarose gel electrophoresis

DNA molecules were separated according to their size on horizontal agarose (Sigma) gels. The percentage of agarose used to make the gel depended on the size range of the DNA molecules to be resolved. Digested genomic DNA or YAC DNA was commonly run on 0.8% agarose gels, whereas smaller fragments, such as plasmid DNA and most PCR products, were run on 1-3% agarose gels. Resolution of polymorphic markers, where the size difference was <50bp and >10bp was performed using 5% Nusieve agarose (Sigma). All agarose gels were made with and run in 0.5 x TBE. To stain the DNA, ethidium bromide was added to agarose gels and the running buffer at a concentration of 250μg/ml of gel/buffer. A tenth of the sample volume of 10x loading buffer was added to each DNA sample prior to loading. Either mini gels (30ml agarose) or midi gels (120ml agarose) were used depending on requirements; gels were run at 25-200V depending on resolution and run-time

required. DNA fragments were visualised on a mid range UV transilluminator and photographed using a video copy processor (Mitsubishi).

### 2.5.3 Polyacrylamide gel electrophoresis

Denaturing polyacrylamide gels were used to resolve DNA sequencing ladders. Glass plates were prepared by thorough washing in detergent and distilled water, followed by 100% ethanol. The front plate was coated with dimethyldichlorosilane (BDH) and left to dry. The plates were then rinsed briefly in 100% ethanol (back plate) or distilled water (front plate) and dried. 0.4mm spacers were placed along the sides of the back plate, the front plate was placed on top, and the sides and bottom of the plates sealed together with tape. Bulldog clips were used to clamp the plates together. Freshly prepared polyacrylamide solution was then poured between the plates and an inverted sharks-tooth comb inserted 3mm into the top of the gel. The gel was allowed to set for at least an hour at room temperature.

Once set the bulldog clips, tape sealing the bottom of the plates and comb were removed. The gel was assembled onto a vertical slab gel apparatus (Gibco, BRL) and prerun in 1 x TBE at 80 volts for one hour. The surface of the gel was then cleaned by flushing with 1x TBE to remove excess urea and gel fragments. The comb was then inserted and the gel loaded with up to 4µl sample. Gels were run at 80 volts for 1 - 3 hours, depending on the size of fragments being resolved.

Following electrophoresis, all tape was removed from the plates and the front plate lifted off, leaving the gel stuck to the back plate. The gel was then transferred onto 3MM Whatman paper, covered with Saran wrap (Dow Chemical Co.) and dried at 80°C on a Drygel slab drier (Hoefer). The Saran wrap was removed and the gel then exposed overnight to Biomax film (Kodak).

Non-denaturing polyacrylamide gels were employed to resolve polymorphic DNA markers with a size difference of 20bp - 2bp. Gels were cast in The Sturdier vertical slab gel apparatus (Model SE 400, Hoefer) using 1.5mm spacers and comb.

Prior to loading, DNA samples, usually 2-10μl of a PCR reaction, were mixed 10:1 with loading buffer. Gels were run in 1 x TBE at 200-300 volts for 2-3 hours.

DNA was visualised by silver staining. The glass plates were separated and the gel, still adhered to the back plate, was washed twice for three minutes in 10% ethanol, 0.5% acetic acid. It was then incubated for 10 minutes at room temperature in 500ml 0.1% $AgNO_3$. The gel was then washed twice for 2 minutes in water and the silver stain developed by incubation in 500ml 0.375M NaOH, 2.6mM $NaBH_4$, 0.15% formaldehyde for 20 minutes. Gels were then fixed in 0.75% $Na_2CO_3$ for 10 minutes and photographed using a video copy processor (Mitsubishi).

### 2.5.4 Pulsed field gel electrophoresis

Plugs were washed once in TE and equilibrated to the running buffer (0.5x TBE) for one hour. A 1% agarose (MP agarose, Boehringer Mannheim) gel in 0.5 x TBE was cast and the plugs inserted into the wells. The wells were sealed with 1% LMP agarose and the gel run in pre-cooled sterile 0.5x TBE at 14°C. Gels were run using the Chef-DR II pulse field tank and control module (Biorad). Run times and conditions varied depending on the size resolution required and were selected according to manufacturers recommendations.

Following electrophoresis, DNA was stained by gently agitating the gel in running buffer containing 250μg/ml ethidium bromide for 20 minutes. The DNA was then visualised and photographed as in section 2.5.2.

## 2.6 Isolation and analysis of yeast artificial chromosomes

### 2.6.1 Media and solutions

AHC broth and agar

AHC is a rich selective medium which lacks uracil and tryptophan. It was used for selective growth of YAC recombinants prior to the isolation of YAC DNA and the production of plugs.

In 1liter water dissolve 1.7g yeast nitrogen base (without amino acids and $NH_4SO_4$) (Difco), 5g $NH_4SO_4$ and 10g casein hydrolysate (low salt). Adjust pH to 5.8. 17-20g of Bacto agar (Difco) were added to each litre of broth to make AHC agar. Autoclave. Add 50ml filter sterilised 40% glucose and 10ml of 2mg/ml adenine sulphate.

### 2.6.2 Screening of the ICRF YAC library by hybridisation

The ICRF YAC library screened consisted of two libraries which had been mechanically gridded at high density onto two filters. The libraries, made by Z. Larin *et al.*, (1991) were constructed from C3H and C57BL/6 DNA respectively and together represent approximately four genome equivalents.

The hybridisation conditions used were as recommended by the ICRF. Briefly filters were prehybridised at 68°C for four hours in; 7% SDS, 0.5M $NaPO_4$ (pH7.2), 1% BSA, 1mM EDTA. Hybridisation was performed using probe labeled as detailed in section 2.11.1. Also included in the hybridisation, to aid in visualisation of background negative clones, was 50ng of AB1380 yeast DNA random prime labeled with $^{35}$S-ATP. Hybridisation was performed at 68°C for 16 hours. Filters were removed from the hybridisation bottles and washed thus; two 5 minute washes in 40mM $NaPO_4$ (pH7.2), 0.1% SDS at room temperature, then two 15 minute washes in 40mM $NaPO_4$ (pH7.2), 0.1% SDS at 65°C.

On receipt of positive clones cells were struck out on AHC plates. Individual colonies were then analysed for marker content by PCR as detailed in section 2.9.1.

### 2.6.3 Screening of YAC libraries by PCR

The St. Marys' Hospital library (Chartier *et al.*, 1992), the Whitehead/MIT library (Kusumi *et al.*, 1993) and the ICRF library (Larin *et al.*, 1991) were all available to be screened by PCR. The library of Chartier *et al.* was constructed from C57BL/10 DNA and represents approximately 3.5 genome equivalents. This was later combined with the ICRF library, constructed from C3H and C57BL/6 DNA and representing approximately four genome equivalents, enabling simultaneous screening of both libraries. (Both libraries were obtained from the HGMP Resource Centre, Hinxton Hall Cambs). The Whitehead/MIT library (purchased from Research Genetics) was constructed from C57BL/6J DNA and represents approximately four genome equivalents.

Both libraries have been arranged in a similar way to facilitate screening by PCR. An initial screen of a primary pool, containing DNA from every YAC in the library identifies a secondary pool. Screening of the secondary pool yields the coordinates identifying an individual clone.

PCR was performed following suppliers recommendations and according to the conditions for each marker set out in section 2.9.1. Following resolution of PCR products by agarose gel electrophoresis the DNA was Southern blotted onto nylon membranes for analysis by hybridisation to appropriate radiolabeled probes. Positive clones were confirmed by PCR analysis as described in section 2.6.2, above.

### 2.6.4 Preparation of YAC DNA in solution

10ml of AHC medium was inoculated with a single colony and grown overnight, shaking at 30°C. 100ml AHC medium was then inoculated with 0.2ml of the overnight culture and grown for 24-48 hours with vigorous shaking, 270rpm, at 30°C, until a cell density of $>10^8$ cells/ml was reached. Cells were harvested by

centrifugation for 10mins at 3000rpm (in a Sorvall T6000) and resuspended in 5ml of 1M sorbitol, 20mM EDTA (pH7.5), 14mM β-mercaptoethanol. Zymolase (ICN Biomedicals) was added to a final concentration of 40μg/ml and the cells incubated for 1 hour at 37°C, with gentle shaking. Cells were then pelleted by centrifugation for 10 mins at 1000rpm and resuspended in 5ml of 4.5M GuHCl, 0.1M EDTA, 0.15M NaCl, 0.05% sarkosyl and incubated at 65°C for 10 minutes. The solution was cooled to room temperature and an equal volume of ethanol added. Nucleic acids were then pelleted by centrifugation for 10 mins at 4000rpm. The pellet was resuspended in 2ml TE (pH 7.4), RNAase added to 100μg/ml and incubated for 30 minutes at 37°C. Proteinase K was then added to a final concentration of 200μg/ml and incubated for 1 hour at 65°C. The DNA was then extracted three times with phenol/chloroform or until no protein was visible at the interface (section 2.3.3). The DNA was then recovered by ethanol precipitation and resuspended in 500μl TE (pH 7.4).

### 2.6.5 Preparation of YAC DNA in agarose plugs

10ml of AHC medium was inoculated with a single colony and grown overnight, shaking at 30°C. 100ml AHC medium was then inoculated with 0.2ml of the overnight culture and grown for 24-48 hours with vigorous shaking, 270rpm, at 30°C, until a cell density of $>10^8$ cells/ml was reached. Cells were harvested by centrifugation for 10mins at 3000rpm (in a Sorvall T6000) and washed once in 50ml 50mM EDTA (pH 8.0). Cells were then resuspended in 2ml 1M sorbitol, 20mM EDTA (pH7.5), 14mM β-mercaptoethanol, 100μg/ml Zymolase (ICN Biomedicals) and warmed to 42°C. An equal volume (2-3ml) of 1% low melting point agarose (Ultrapure LMP agarose, Gibco BRL) in 1M sorbitol, 20mM EDTA, 14mM β-mercaptoethanol, also at 42°C, was then added. After gentle but thorough mixing the cell suspension was then dispensed, in 100μl aliquots, into plug molds (plug molds were wells of a 96 well plate, the plastic bottoms of which had been

removed. Wells were thoroughly washed and sealed with tape). The plugs were placed on ice until set and then expelled from the mold into 25ml 1M sorbitol, 20mM EDTA, 14mM β-mercaptoethanol, 10mM Tris.HCl (pH7.5), 100μg/ml Zymolase and incubated at 37°C with gentle shaking. The solution was then replaced with 25ml yeast lysis solution (1% SDS, 100mM EDTA, 10mM Tris.HCl pH 8.0) and incubated for 1 hour at 37°C with gentle shaking. The solution was then replaced with fresh yeast lysis solution and the plugs incubated overnight hour at $37^0$C with gentle shaking. Plugs were stored at 4°C in 0.5M EDTA (pH 8.0).

### 2.6.6 Partial digestion of YAC DNA in agarose plugs

Plugs, stored in 0.5M EDTA (pH 8.0) were washed by gentle agitation in 100 volumes of TE at room temperature for 15 mins. This was repeated six times. For each digestion 30μl, or a third of a plug (containing approximately 2μg DNA) was used. Plugs slices were soaked in 1ml of the recommended 1x restriction enzyme buffer, supplemented with 200μg/ml BSA for 1 hour at room temperature, with frequent inverting. The volume was then reduced to 100μl. For each enzyme three plug slices were digested with 30 units, 10 units and 1 unit (except *BssH*II and *Sfi*I which required 10 units, 1 unit and 0.1 unit) for 1 hour exactly at the recommended temperature. Following digestion the reaction solution was removed and the plug rinsed in 1ml ice-cold TE. Plugs were then kept on ice in 1ml TE prior to gel loading.

### 2.6.7 Isolation of YAC end clones by inverse PCR

Adapted from Arveiler and Porteous, (1991).

500ng of total YAC DNA was digested to completion with *EcoR*V, *Rsa*I and *Taq*I, for the isolation of left arm clones and *EcoR*V, *Hind*III, *Pst*I and *Rsa*I, for the isolation of right arm clones. The digestion products were purified using Geneclean and ligated at a concentration of 1-2.5ng/μl. At this DNA concentration intra-

molecular ligation events are favored over inter-molecular events resulting in the circularisation of DNA fragments.

1μl of the ligation reaction was then used to amplify end clone fragments by PCR using primers 372/373 for left end clones and 374/556 for right end clones, except for *Hind*III digested DNA which required primers 375/556 (see table 2.1). For reactions giving a very low yield of product or the amplification of multiple fragments, a nested PCR was performed using C626/372 for left end clones and C627/374 or for right end clones, except for *Hind*III digested DNA which required primers C627/375.

End clone fragments were isolated following separation by agarose gel electrophoresis.

## 2.7 Isolation and analysis of lambda clones

The mouse genomic lambda library, purchased from Stratagene, was made from female liver of a 129/Sv mouse in the vector lambda FixII.

### 2.7.1 Plating out the lambda library

**(a)** Preparation of plating bacteria.

Plating bacteria, strain XL1-Blue MRF (Stratagene), were grown overnight, shaking at 250rpm, 37°C, from a single colony in 50ml LB medium supplemented with 0.2% maltose. Cells were harvested by centrifugation for 10min at 4000rpm, room temperature (in a Sorval T6000) and resuspended in 20ml 10mM $MgSO_4$. Cells were stored at 4°C and were viable for 3 weeks.

**(b)** Determination of phage titre.

Tenfold serial dilutions of bacteriophage in SM buffer (0.58% NaCl, 0.2% $MgSO_4$, 50mM Tris.HCl, (pH 7.5), 0.01% gelatin) were prepared to cover the range 1 - $10^5$ p.f.u./ml. 0.1ml bacteriophage were then added to 0.1ml plating bacteria and incubated for 20 minutes at 37°C. 3ml of top agar (0.7% agar (Sigma) in LB), at 47°C,

was then added, mixed and poured onto 90mm LB plates. Once set the plates were inverted and incubated at 37°C overnight. The number of plaques produced by each bacteriophage dilution was then used to determine the titre of the original stock.

**(c)** Plating the library.

The library was plated immediately after determining the bacteriophage titre using the same preparation of plating bacteria. 3.3ml of plating bacteria were incubated at 37°C for 20 minutes with the quantity of phage, in 3.3ml SM buffer, estimated to give 1,100,000 plaques. 71.5ml top agar at 47°C was then added and mixed well. 7.1ml was then pipetted onto each of ten prewarmed (37°C) 150mm LB plates. When set the plates were inverted and incubated at 37°C for 6-12 hours, until plaques were almost confluent. Plating at this density; ten plates of 100,000 plaques represents approximately five genome equivalents.

**(d)** Taking lifts

Plates were cooled for 1 hour at 4°C to harden the top agar and 139mm Hybond-N nylon membrane (Amersham) placed on the surface of each plate. For orientation a syringe needle charged with Indian ink was stabbed through each filter and plate in four asymmetric locations. After 1 minute the membrane was peeled off and placed face up for 5min on 3MM Whatman paper soaked in denaturing solution (1.5M NaCl, 0.5M NaOH). The membrane was neutralised on filter paper soaked in 1.5M NaCl, 0.5M Tris.HCl (pH 7.2), 1mM EDTA for 5min. The membrane was then washed in a solution of 2 x SSC for 30sec, blotted dry and the DNA crosslinked to the membrane using U.V. light (UV Stratalinker, Stratagene). Replica membranes were prepared in a similar fashion by placing the membrane on the agar surface for 2 minutes and marking the membrane through the same holes as the original membrane.

### 2.7.2 Library screening

$^{32}P$ radiolabeled probes were hybridised to the nylon membranes as described in section 2.12. Positive plaques, as identified by signals produced on both

membranes taken from each plate, were picked, using a pasteur pipette, into 1ml SM buffer containing one drop of chloroform. After incubation at room temperature for 2 hours the phage suspension was diluted by tenfold serial dilution. Dilutions of $10^{-3}$ to $10^{-6}$ were used to infect plating bacteria and plated out onto 90mm LB plates, as described in section 2.7.1 (b). Plates were then chosen which had a plaque density of 10 - 500 and the phage lifted onto 79mm nylon membranes (Amersham), as described above in 2.7.1 (d) (except replicas were not taken). These membranes were then hybridised with the same probe used for the initial screening and positive plaques isolated as above. This process was repeated until a single positive plaque could be unequivocally picked, without risk of contamination from neighbouring plaques. Positive plaques were stored in 1ml SM buffer, containing one drop of chloroform at 4°C.

### 2.7.3 Purification of lambda DNA

Lambda DNA was purified from plate lysates using a lambda midiprep kit (Qiagen), according to manufacturers instructions.

## 2.8 Oligonucleotides

### 2.8.1 Oligonucleotide Synthesis

Oligonucleotides were synthesised (by Agnes Gallagher) as ammonium stocks on an Applied Biosystems 381A oligonucleotide synthesiser.

Oligonucleotides were precipitated from ammonium stocks by ethanol precipitation of 200µl of the stock. The precipitated DNA was resuspended in 200µl water and the concentration assessed by measuring the $Absorbance_{260nm}$. An absorbance of 1 corresponds to a concentration of 25µg/ml for single stranded DNA.

### 2.8.2 PCR primer Design

Primers were designed using the programme Oligo4 (Hybaid). Primers were designed to be between 19 and 30 nucleotides in length. They were chosen to be stable oligonucleotides with a $T_m$ between 55°C and 75°C. A maximum hairpin structure or potential dimerism of 3 nucleotides (preferably less) was permissible, but only a 2 nucleotide internal match if it involved the 3′ end. If possible the A+T : C+G ratio was near 50%.

### 2.8.3 Duplexing oligonucleotides

Preparation of DNA linkers involved the formation of a duplex between two complementary (or partially complementary) oligonucleotides. Equimolar quantities of the two oligonucleotides, typically 150μg/ml each, were mixed together in 10μl 1x TM (10mM Tris.HCl, (pH 7.5), 5mM $MgCl_2$) and the tubes placed in a thermal cycler at 99°C. The thermal cycler was then switched off and allowed to cool to below 30°C over a period of 2 hours.

## 2.9 Amplification of DNA by the polymerase chain reaction

### 2.9.1 PCR conditions

PCR reactions were routinely performed in a volume of 20-50μl in 0.8ml thin walled microcentrifuge tubes (Robbins Scientific). Reaction components consisted of 1μM of each primer, 200μM each dNTP (Advanced Biotechnologies), 1-2 units Amplitaq (Perkin Elmer Cetus) in 1x reaction buffer (10mM Tris.HCl, (pH 8.3), 50mM KCl, 1.5mM $MgCl_2$, 0.01% (w/v) gelatin) and 0.1% Triton X-100 (Sigma). The quantities of DNA template commonly used were; 50-100ng of mouse genomic DNA, 10ng of YAC or lambda DNA and 0.1ng of plasmid DNA. Amplification directly from bacterial or yeast cells was performed by the addition of a very small

number of cells into the reaction mix using a disposable tip. The reaction was overlaid with 40μl mineral oil (Sigma).

PCR programmes were run on a Hybaid Omnigene thermal cycler. The basic programme was 35 cycles of 3 different steps:

(a) Denaturation; A short step at 93-94°C. Usually 3 minutes in the first cycle and 30 seconds thereafter. Amplification from intact bacterial or yeast cells was performed with an initial denaturation step of 95°C for 5 minutes.

(b) Annealing; The annealing temperature is critical in determining the success of the PCR reaction, and is dependent on the structure of the primer. Where possible primer pairs of approximately equal $T_m$ were used. Annealing temperatures were routinely 5°C lower than the $T_m$ of the primers used, calculated according to the equation; $T_m = 4(G+C) + 2(A+T)$. Annealing steps were 30-60 seconds.

(c) Extension; Carried out at 72°C. Extension times were routinely 30-60 seconds. Where product of over 1kb was predicted the extension time was increased according to the estimation; 1kb of sequence is synthesised per minute of extension time.

Following PCR, 5-10μl of each reaction was analysed by electrophoresis on 1-5% agarose gels.

**Steps to optimise reaction conditions:**

Empirical optimisation of reaction conditions was sometimes necessary to achieve efficient reactions. This was achieved primarily by variation in annealing temperature.

'Hot Start' PCR (Chou *et al*, 1992) was commonly used. The PCR reaction was set up without enzyme. The reaction was heated to 90°C for 3 minutes and the enzyme then added. Addition of the enzyme to the DNA at a temperature at which the DNA is denatured helps to prevent mispriming before the initial denaturing temperature is reached.

For some applications (chapter 4, 4.2.1) Taq Extender (Stratagene) was used, according to manufacturers instructions, to improve reaction efficiency.

Except for the PCR primers detailed in other chapters, sequences of all primers and specific reaction conditions used are listed in table 2.1.

### 2.9.2 Reverse transcription PCR

The following components were assembled in a total volume of 20μl; 1μg of RNA in water (previously heated to 95°C for 10 mins, and cooled on ice), 1mM each dNTP (Pharmacia), 1unit/μl RNAsin (Boehringer Mannheim), 100pmol random hexamer (Pharmacia), 1μg BSA (Boehringer Mannheim), 1 x PCR buffer (Perkin Elmer Cetus) and 200units Moloney murine leukaemia virus reverse transcriptase (Gibco BRL). The reaction was incubated for 10 minutes at 23°C and then for 1 hour at 42°C. The reaction was then incubated at 95°C for 5 minutes and then chilled on ice.

The 20μl reaction was then used directly in a 100μl PCR reaction following the parameters outlined in section 2.9.1.

### 2.9.3 Catch linker PCR

Catch linker PCR is a method for generating a simplified resource of cloned genomic DNA. DNA is digested, usually with *Sau*3AI, and then a short double stranded 'catch linker' is ligated onto each cohesive end. The intervening DNA can then be amplified by PCR using primers designed to the catch linkers.

For the preparation of catch linkered YAC DNA the YAC was isolated from yeast chromosomes by pulsed field gel electrophoresis. Lambda DNA was catch linkered without prior removal of the vector arms. 1μg of DNA was digested with *Sau*3AI at 37°C for 1 hr. Digestion was checked by analysis of a small aliquot on a

**Table 2.1 Oligonucleotides used for PCR amplification**

For *Gli2* and *Inhβb* primers, F designates the forward primer and R the reverse primer. The number indicates the position of each primer relative to the published sequence (for *Gli2* see Hughes *et al.*, 1997, and for *Inhβb* see chapter 4).

| Name | Application | Sequence 5′→ 3′ | Reaction conditions |
|---|---|---|---|
| ***Gli2* primers** | | | |
| I51<br>I52 | *Gli2* F395<br>*Gli2* R541 | TTGATACGACTTTCTCCACACCC<br>GAGGCCCCTGGCGGCAGAGATC | 95°C for 3 min x 1<br>94°C for 45 sec<br>65°C for 30 sec x 35<br>72°C for 30 sec |
| 511<br>512 | *Gli2* F1583<br>*Gli2* R1733 | AAGGCCTACTCCCGCCTGGAGAACC<br>TCTCATTGGAGTGAGTGCGGTTCTG | 95°C for 3 min x 1<br>94°C for 45 sec<br>65°C for 30 sec x 35<br>72°C for 30 sec |
| I35<br>I34 | *Gli2* F4531<br>*Gli2* R4651 | GGATGATGGTGATCACTCGAG<br>GATGGAGGGCAGTGTCAAGG | 95°C for 3 min x 1<br>94°C for 45 sec<br>65°C for 30 sec x 35<br>72°C for 30 sec |
| F264<br>F265 | *Gli2* F2840<br>*Gli2* R2980 | GGTGCCTGCCCACATCCACTGG<br>GTCGCTGGCCCGCCGTGTGCTG | 95°C for 3 min x 1<br>94°C for 45 sec<br>68°C for 30 sec x 40<br>72°C for 90 sec |
| 1a<br>1b | *Gli2* F80<br>*Gli2* R1129 | GTCCGCCACCAAAGAGTTTGAGCC<br>AGTTGGGTAGGCATGGTGCTGA | 95°C for 3 min x 1<br>94°C for 45 sec<br>65°C for 30 sec x 40<br>72°C for 60 sec |
| 5a<br>5b | *Gli2* F3817<br>*Gli2* R4781 | GCTGAGCCCAAACATTGTCAGCGG<br>GTGTTGCTCAGGTCTCCTGTGCCA | 95°C for 3 min x 1<br>94°C for 45 sec<br>68°C for 30 sec x 35<br>72°C for 60 sec |
| D350<br>D514 | *Gli2* F1289<br>*Gli2* R1881 | CTCAAGGAAGACCTGGACAGGGATGA<br>TGAGCAGTGGAGCACGGACATGCAC | 95°C for 3 min x 1<br>94°C for 45 sec<br>65°C for 30 sec x 35<br>72°C for 30 sec |
| 12a<br>1b | *Gli2* F841<br>*Gli2* R1129 | GATTCGGACCTCTCCCAACTCGCTGG<br>AGTTGGGTAGGCATGGTGCTGA | 95°C for 3 min x 1<br>94°C for 45 sec<br>65°C for 30 sec x 35<br>72°C for 30 sec |

| Name | Application | Sequence 5′ - 3′ | Reaction conditions |
|---|---|---|---|
| ***Inhβb* primers** | | | |
| I181<br>I182 | *Inhβb* F492<br>*Inhβb* R1162 | CTTCTTCGTCTCTAATGAAGG<br>GAGCATGGACATGGAGCTCAGC | 95°C for 3 min x 1<br>94°C for 60 sec<br>58°C for 30 sec x 35<br>72°C for 60 sec |
| I734<br>I735 | *Inhβb* microsatelite marker | GGAAGAGTGGTCTAGTTTCTATGG<br>AGCCTGGAGGCTTAACTAGCAC | 95°C for 3 min x 1<br>94°C for 60 sec<br>65°C for 30 sec x 35<br>72°C for 30 sec |
| *En1* primers | | | |
| 257<br>829 | *En1* homeobox | GAAGCTAAAGAAGAAAAAGAACG<br>TCTCGTCTTTGTCCTGAACC | 95°C for 3 min x 1<br>94°C for 60 sec<br>55°C for 30 sec x 35<br>72°C for 60 sec |
| 989<br>990 | *En1* homeobox | TCCGTCCTCTGGTCCACGC<br>TCTCGTCTTTGTCCTGAACC | 95°C for 3 min x 1<br>94°C for 60 sec<br>60°C for 30 sec x 35<br>72°C for 300 sec |
| *En2* primers | | | |
| D352<br>D353 | *En2* homeobox | CACAGGAGCTACGCCTGAACGAGTCT<br>CTACTCGCTGTCCGACTTGCCCTCCT | 95°C for 3 min x 1<br>94°C for 60 sec<br>68°C for 30 sec x 40<br>72°C for 60 sec |
| D448<br>D447 | *En2* homeobox | TGGCCCGCTTGGGTCTACTGCAC<br>TTGTACAGGCCCTGTGCCATGAG | 95°C for 3 min x 1<br>94°C for 60 sec<br>68°C for 30 sec x 40<br>72°C for 60 sec |
| F886<br>F887 | *En2* 3′ untranslated polymorphism | CACTAAACAGAAGACATTGTGGAG<br>CTCCCGAATCTAGAGGACATGTG | 95°C for 3 min x 1<br>94°C for 45 sec<br>62°C for 30 sec x 35<br>72°C for 30 sec |
| *Pmsc2* primers | | | |
| F1.1<br>R5.5 | *Pmsc2* polymorphic marker | CAAGGCAGATCAAGCAGGTT<br>TGCAGCCAAATCCCAGAGG | 95°C for 3 min x 1<br>94°C for 60 sec<br>55°C for 30 sec x 40<br>72°C for 2 min |
| F4<br>4R2 | *Pmsc2* polymorphic marker | GGAGCTCGAATGGTTCGTGAG<br>TACTGTTTGGACACAGGCGGG | 95°C for 3 min x 1<br>94°C for 45 sec<br>60°C for 30 sec x 35<br>72°C for 90 sec |
| *Emv1* primers | | | |
| G313<br>G454 | *Emv1* polymorphic marker | CTCAGTGTTGTCCTGATACGTGAC<br>CCATTCCTCACGGAGCCAAGTG | 95°C for 3 min x 1<br>94°C for 45 sec<br>65°C for 30 sec x 35<br>72°C for 30 sec |

| Name | Application | Sequence 5′ - 3′ | Reaction conditions |
|---|---|---|---|
| YAC end clone markers | | | |
| H58<br>H59 | 129B7RE | GACTGGATGTTCCCACTGAGGAG<br>GCAAACTATGACCTTCCTTCTGGG | 95°C for 3 min x 1<br>94°C for 45 sec<br>70°C for 30 sec x 35<br>72°C for 90 sec |
| I580<br>I581 | 67G3LE | GCATGGGTGGTGGGGGTCTGA<br>AGACAGGAGAAGCCATTTCAGGG | 95°C for 3 min x 1<br>94°C for 45 sec<br>65°C for 30 sec x 35<br>72°C for 30 sec |
| I975<br>I976 | 119A4RE | GGGCAGCCATTAGTCATAGCTG<br>GGCTCTTAGGCATGGGTTTCAC | 95°C for 3 min x 1<br>94°C for 45 sec<br>63°C for 30 sec x 35<br>72°C for 30 sec |
| K680<br>K682 | 89H7RE | TAATTGCTGGCTGCTCTTATCG<br>GGAAGACAGACACAGCTCCTG | 95°C for 3 min x 1<br>94°C for 45 sec<br>60°C for 30 sec x 35<br>72°C for 30 sec |
| J141<br>J142 | 76B3LE | CAGCTCCTGGGGTTGGTAGTT<br>CACAGAGAAACCCTGTCTCGA | 95°C for 3 min x 1<br>94°C for 45 sec<br>60°C for 30 sec x 35<br>72°C for 60 sec |
| J145<br>J146 | 76B3RE | CTGCAGATGACAGAAACAAGCG<br>GTCCTATCCTGGTTCCTCTTG | 95°C for 3 min x 1<br>94°C for 45 sec<br>58°C for 30 sec x 35<br>72°C for 30 sec |
| H358<br>H360 | 114C12LE | AAGGCCCTATCAGCCTCCCAG<br>GAGAGGCTGTGAATTCCAGAGAC | 95°C for 3 min x 1<br>94°C for 45 sec<br>65°C for 30 sec x 35<br>72°C for 30 sec |
| J114<br>J115 | 114C12RE | CTATCTGTGGGACATTGTCTTG<br>CCGACTCTTCGCCTGCTTTC | 95°C for 3 min x 1<br>94°C for 45 sec<br>58°C for 30 sec x 35<br>72°C for 60 sec |
| J143<br>J144 | 130G11LE | AAACGTCCTCTACACCTTAAAC<br>ATGCTTTCTCTGCATCTAACGAG | 95°C for 3 min x 1<br>94°C for 45 sec<br>58°C for 30 sec x 35<br>72°C for 60 sec |
| I143<br>I144 | 130G11RE | GTAGGCCTCTGTAGAGTTTCTCGG<br>TGCCCCATAGGGTGGGCTGATG | 95°C for 3 min x 1<br>94°C for 45 sec<br>68°C for 30 sec x 35<br>72°C for 60 sec |
| I261<br>I262 | 175D3LE | AGAAGAAGTCTTTGCCACAAAT<br>AACCCTTATACTGTTTCTTATCA | 95°C for 3 min x 1<br>94°C for 45 sec<br>55°C for 30 sec x 35<br>72°C for 30 sec |
| D965<br>D966 | D9P8RE | CACTGCATTCACATACACATACC<br>AAAGTTAGATGATAGACCAGAAC | 95°C for 3 min x 1<br>94°C for 45 sec<br>60°C for 30 sec x 35<br>72°C for 60 sec |

| Name | Application | Sequence 5′ - 3′ | Reaction conditions |
|---|---|---|---|
| G807<br>G808 | CO22RE | GATTACAGGCATGAGCCACCAC<br>GGAGTTCCCTCTAATGAGGCAG | 95°C for 3 min x 1<br>94°C for 45 sec<br>65°C for 30 sec x 35<br>72°C for 60 sec |
| F920<br>F921 | 98D8LE | GAACAGCAATGGTTTGTGCTG<br>GATTTTCTGGAGTCCACCTTC | 95°C for 3 min x 1<br>94°C for 45 sec<br>55°C for 30 sec x 35<br>72°C for 30 sec |
| DMit microsatelite markers | | | |
| MIT microsatellite markers were purchased from Research Genetics and used according to manufacturers recomendations. Primer sequences are listed at http://www-genome.wi.mit.edu/. | | | |
| YAC end cloning | | | |
| 372<br>373 | Left end clone | GAATTGATCCACAGGACGGG<br>GCCAAGTTGGTTTAAGGCGC | 95°C for 3 min x 1<br>94°C for 30 sec<br>60°C for 30 sec x 35<br>72°C for 2.5 min |
| 374<br>556 | Right end clone | GGAAGAACGAAGGAAGGAGC<br>GCCCGATCTCAAGATTACG | 95°C for 3 min x 1<br>94°C for 30 sec<br>60°C for 30 sec x 35<br>72°C for 2.5 min |
| 375<br>556 | Right end clones (for *Hind*III digests) | AAACTCAACGAGCTGGACGC<br>GCCCGATCTCAAGATTACG | 95°C for 3 min x 1<br>94°C for 30 sec<br>60°C for 30 sec x 35<br>72°C for 2.5 min |
| 372<br>C626 | Left end clone, nested PCR | GAATTGATCCACAGGACGGG<br>AATTTATCACTACGGAATTC | 95°C for 3 min x 1<br>94°C for 30 sec<br>50°C for 30 sec x 35<br>72°C for 2.5 min |
| 374 or 375<br>C627 | Right end clone, nested PCR | see above<br>GATCTCAAGATTACGGAATTC | 95°C for 3 min x 1<br>94°C for 30 sec<br>55°C for 30 sec x 35<br>72°C for 2.5 min |

| Name | Application | Sequence 5′ - 3′ | Reaction conditions |
|---|---|---|---|
| Vectorette PCR, linkers and vector primers | | | |
| | vectorette linker: upper strand | GATCAAGGAGAGGACGCTGTCTGTCG AAGGTAAGGAACGGACGAGAGAAGGG AGAG | 95°C for 3 min x 1<br>94°C for 60 sec<br>*50-68°C for 45 sec x 35<br>72°C for 2.5 min |
| | lower strand | TTCCTCTCCTGTCGCTAAGAGCATGC TTGCCAATGCTAAGCTCTTCCCTCT C | |
| G400 | vectorette primer (from Riley et al., 1990) | CGAATCGTAACCGTTCGTACGAGAA CGCT | *annealing temperature depends on the TM of the second primer. |
| 963L<br>962L | *Sau*3AI linker | GATCGTCGACGGTACCGAATTCT GTCAAGAATTCGGTACCGTCGAC | For linker PCR using 962L<br>95°C for 3 min x 1<br>94°C for 60 sec<br>55°C for 45 sec x 35<br>72°C for 1.5 min |
| 963L<br>964L | Blunt end linker | GATCGTCGACGGTACCGAATTCT GTCAAGAATTCGGTACCGTCGACGA TC | For linker PCR using 962L<br>95°C for 3 min x 1<br>94°C for 60 sec<br>55°C for 45 sec x 35<br>72°C for 1.5 min |
| T3<br>T7 | Amplification of pBluescript inserts | CGACTCACTATAGGGCGAATTGG CAATTAACCCTCACTAAAGGGAAC | 95°C for 3 min x 1<br>94°C for 45 sec<br>65°C for 30 sec x 35<br>72°C for 1 min |

3% agarose gel. Fully digested DNA was seen as a smear centered around 300-400bp. DNA was then purified using the Geneclean procedure (2.3.3).

Approximately 20ng of DNA was then ligated, in a volume of 20μl, to 100ng of the linker 962L/963L (see table 2.1), prepared as in section 2.8.3. This represents an estimated 5-fold molar excess of linker to DNA.

80μl of water was then added to the ligation reaction and 1μl amplified by PCR, using oligonucleotide 962L as the primer. The PCR programme used was: denaturation: 95°C for 3min (1 cycle), denaturation, 94°C for 60sec; annealing, 55°C for 45sec; extension, 72°C for 1min (35 cycles).

5μl of the PCR product was examined on a 1.5% agarose gel. The remainder was ethanol precipitated and resuspended as appropriate to future use.

## 2.10 Transfer of DNA to membranes

### 2.10.1 Alkali transfer of DNA

DNA was transferred from gels to nylon membranes by capillary blotting. This method was adapted from Southern (1975). Gels containing DNA fragments greater than 15kb were initially treated in 0.25M HCl at room temperature for 20 mins with gentle agitation and then rinsed in deionised water. DNA was denatured by gently shaking of the gel in denaturing solution (0.5M NaOH, 1.5M NaCl) for 30 minutes. The gel was then equilibrated in alkali transfer buffer (0.4M NaOH, 1.0M NaCl) with gentle agitation for 30 minutes A large strip of 3MM filter paper (Whatman) was soaked in alkali transfer buffer and placed on a board. The ends of the paper were placed in a reservoir of alkali transfer buffer, forming a wick. The gel was placed, inverted, on top of the wet filter paper, then a piece of nylon membrane (Hybond-N, Amersham) cut to be just larger than the gel was placed directly onto the gel, ensuring no air bubbles were trapped. Two pieces of 3MM blotting paper, pre-soaked in alkali transfer buffer, were then placed on top of the membrane and air

bubbles carefully removed. Any exposed wick was screened off with Saran wrap (Dow Chemical Company), and a weighted stack of paper towels placed on top of the gel.

Gels were blotted for either 5 hours or overnight. After blotting, the membranes were rinsed in 2 x SSC (20 x SSC; 3M NaCl; 0.3M $Na_3C_6H_5O_2.2H_2O$, pH7.0) and left to air dry. DNA was then covalently linked to the filter by exposing it to 1200μJoules of UV irradiation in a Stratalinker (Stratagene). Nylon membranes were stored in Saran wrap or between two sheets of 3MM paper at room temperature.

## 2.11 Radiolabelling of DNA

### 2.11.1 Random prime labeling of DNA probes

Double stranded DNA was radiolabeled with [α-$^{32}$P]-dCTP using the Random Primed DNA labeling kit (Boehringer Mannheim) according to manufacturers instructions (adapted from Feinberg and Vogelstein, 1984).

Purification of labeled probe from unincorporated [α-$^{32}$P]-dCTP was performed using Sephadex G-50 Nick columns (Pharmacia Biotech) according to manufacturers instructions. This also gave an estimation of the percentage incorporation of radiolabel by comparing the c.p.s. of the isolated probe to that of the [α-$^{32}$P]-dCTP remaining in the column.

### 2.11.2 Preannealing of repetitive sequences

Background hybridisation due to repetitive elements within probes was suppressed by preannealing probes with repeat containing DNA. The probe was denatured at 100°C for 10 mins in the presence of 1mg sonicated mouse genomic DNA. The two were then incubated in a pre-warmed lead pot at 68°C for 30 minutes before addition to the hybridisation mixture.

### 2.11.3 End-labeling of oligonucleotides

This labeling reaction involves the transfer of the radiolabeled terminal phosphate group of [$\gamma$-$^{32}$P]ATP to the terminal 5'-OH group of the oligonucleotide. 50-100ng oligonucleotide DNA was mixed with 3$\mu$l (30$\mu$Ci) [$\gamma$-$^{32}$P]ATP (Amersham) and 10 units PNK (Boehringer Mannheim), in a total of 20$\mu$l 1x PNK buffer (5mM Tris.HCl (pH 8.0); 1mM $MgCl_2$; 0.5mM DTT). The reaction was incubated at 37°C for 30mins. The probe was added directly to the prehybridisation solution.

## 2.12 Hybridisation protocols

### 2.12.1 Solutions

Hybridisation solution

5x SSPE (20 x SSPE: 3.6M NaCl; 0.2M $NaH_2PO_4.H_20$; 0.02M EDTA (pH 7.4); 0.5% SDS and 5x Denhardts' solution (100x Denhardts' solution: 2% ficoll (Sigma); 2% PVP (Sigma); and 2% BSA (Sigma). Filter before use and store at -20°C).

Oligonucleotide hybridisation solution

5x SSC; 0.05% BSA; 0.05% ficoll; 0.1% SDS; 0.05% PVP; 0.1% sodium pyrophosphate. Filter before use.

### 2.12.2 Prehybridisation of filters

Nylon membranes were placed between slightly larger sheets of hybridisation gauze (Hybaid), rolled up and inserted into the appropriate sized glass hybridisation bottle (Hybaid). Up to 4 membranes could be placed in the same bottle.

Sonicated salmon sperm DNA, denatured by boiling for 10 minutes, was added to prehybridisation solution at a final concentration of 100$\mu$g/ml. Sufficient hybridisation solution to just cover the membranes, typically 10-30ml, was then added and the bottle rotated in a hybridisation oven (Hybaid), at the temperature to

be used for hybridisation, for a minimum of 1 hour. In order to expel air bubbles from the membranes and gauze the bottle was initially rotated in the direction such that the membranes and gauze would roll inwards until they had formed a tight roll. The bottle was then turned around so that the membranes unrolled as they rotated.

### 2.12.3 Hybridisation and washing conditions

For random-primed probes, hybridisation was carried out at 68°C overnight. For oligonucleotide probes, hybridisation was performed overnight at a temperature 5°C lower than the $T_m$ of the oligonucleotide (approximately 50°C). $T_m$ was calculated according to $T_m$ (°C) = 2 x (A +T) + 4 x (G+C).

Filters were washed in the hybridisation bottles using wash solutions of increasing stringency. Initially filters were rinsed twice in 2 x SSPE, 0.1% SDS at 68°C. Then bottles were half filled with 1 x SSPE, 0.1% SDS and rotated at 68°C for 15 minutes. This was then replaced with a wash solution of 0.1 x SSPE, 0.1% SDS and rotated at 68°C for a further 20 minutes. For oligonucleotide probes, the washing steps were far less stringent; 2x 10 minutes in 2x SSC, 0.1% SDS at hybridisation temperature. The filters were then blotted on 3MM paper and wrapped in Saran wrap (Dow Chemical Company).

### 2.12.4 Detection of hybridisation signal

Autoradiography

Filters were placed in a light-tight cassette with a signal intensifying screen. They were then exposed to X-OMAT x-ray film (Kodak) for a length of time dependent on the strength of signal, as judged by measurement of c.p.m. using a Geiger counter (several minutes to several days). Filters hybridised to $^{32}P$-labeled probes were exposed at -70°C, those labeled with $^{35}S$ were exposed at room

temperature. Stratagene Glogos II luminescent markers were used for alignment. The film was developed on an automatic x-ray film processor RGII (Fuji).

Phosphorimaging

Alternatively, the filters were exposed to a phosphor screen (Molecular Dynamics) for hours to several days. The screen was then scanned on a PhosphorImager (Molecular Dynamics), where a laser beam converts the radioactive signal into a digital image, with variations in the pixel value proportional to the amount of radioactive signal present. The grey-scale image was adjusted as desired and then printed on a laser printer.

### 2.12.5 Removal of probe from filters

500ml of a 0.1% SDS solution was heated in a microwave oven until boiling and then poured over the membranes. The solution was then left to cool to room temperature, with gentle agitation. If necessary this procedure could be repeated. Care was taken to prevent membranes from fully drying out as this prevents successful removal of probe. Membranes were then exposed to x-ray film overnight to check for complete probe removal.

## 2.13 Sequencing of DNA

### 2.13.1 Sequencing of PCR products

This method has been modified from Winship, (1989). PCR product of interest was isolated from background product by agarose gel electrophoresis followed by Geneclean purification (2.3.3). DNA concentration was checked by running 1µl on an agarose gel next to a known amount of size marker.

The sequencing reactions were carried out using a Sequenase Version 2.0 DNA sequencing kit (Amersham). 50 - 100ng of DNA was mixed with 150ng of the appropriate sequencing primer, 1.25µl DMSO (Sigma) and 2.5µl 5x Sequenase

reaction buffer and made up to a total volume of 12.5μl with water. The mix was then denatured in a boiling water bath for 3 mins. The tube was then immediately plunged into liquid nitrogen.

A labeling mix was made up from the kit containing 1μl DTT, 2μl dGTP labeling mix (diluted 1:5 with water), 0.5μl of 10μCi/μl [α-$^{35}$S]-dATP (Amersham) and 2μl Sequenase enzyme (diluted 1:8 in enzyme dilution buffer). One tube of DNA/primer mix was then thawed quickly by rubbing between the fingertips, and 5.5μl of the labeling mix immediately added, mixed by pipetting and incubated for exactly one minute at room temperature. 3.5μl of this was then immediately aliquoted into 4 wells of a 96 well plate which had been prewarmed to 37°C. Each well contained 2.5μl of one of the 4 termination mixes with 10% DMSO. The termination reactions were incubated for 5 minutes and the reaction stopped by the addition of 4μl kit stop mix (95% formamide; 20mM EDTA; 0.05% bromophenol blue; 0.05% xylene cyanol FF). Reactions were stored at -20°C until loading on a gel. Prior to loading they were heated to 80°C for 5 minutes.

### 2.13.2 Sequencing of plasmid inserts

Double stranded plasmid DNA was sequenced using the Sequenase Version 2.0 DNA sequencing kit (Amersham), employing the alkaline denaturation method according to the manufacturers instructions

### 2.13.3 Analysis of sequence data

Sequences obtained were edited and manipulated using GCG (Genetics Computer Group) programmes available through the HGMP Computing service (Devereux *et al*, 1984). Database searches were performed using the GCG Blast (Altschul *et al*, 1990) programme.

## 2.14 Calculation of genetic distance

Genetic distances, measured in centimorgans, cM, were calculated using the following equation:

$$\text{Distance in cM} = (R/N) \times 100$$

where R = the number of recombinants

N = the number of animals haplotyped.

Standard error was calculated using the equation:

$$SE = \frac{\sqrt{pq}}{\sqrt{n}} \times 100$$

where p = recombination frequency

q = 1 - p

n = number of animals

# CHAPTER 3

## GENETIC AND PHYSICAL CHARACTERISATION OF THE *Dominant hemimelia* REGION ON CHROMOSOME 1

### 3.1.1 Introduction

The mutation *Dominant hemimelia*, *Dh*, has previously been genetically defined to a region on the central portion of chromosome 1. Studies in this laboratory have shown the closest flanking markers to *Dh* to be the 3′ untranslated region of *Gli2* (0.5±0.3cM) proximally and *En1* (0.7±0.4cM) distally (Higgins *et al.*, 1992, Hughes *et al.*, 1997). As part of a positional cloning strategy aimed at identifying the molecular basis of the *Dh* mutation the work presented in this chapter describes the physical characterisation of this genetic interval. A long range genomic restriction map covering 2.4Mb has been generated and a YAC contig spanning much of this region assembled. The genetic and physical maps are integrated by the production of markers enabling the placement of loci on both maps. The genetic and physical location of two genes, *Gli2*, a member of the *Gli* family of zinc finger genes, and *Inhibinβb*, a member of the *TGFβ* superfamily, within the *Dh* critical region are described.

### 3.1.2 *Dh* maps in the vicinity but is not allelic to *En1* on chromosome 1

The map position of *Dh*, on the central portion of chromosome 1, became of interest when it was established by recombinant inbred strain analysis that the homeobox containing gene *En1* mapped close to or at the *Dh* locus (Hill *et al.*, 1987, Joyner and Martin, 1987). *En1* is a mammalian homologue of the *Drosophila* segment polarity gene *engrailed* which is known to play a vital role in the control of body plan segmentation during embryogenesis. Also since homeobox containing genes are known to play key roles in the control of mammalian development a developmental role for *En1* was assumed. The observation that *En1* mapped near the

developmental mutation *Dh* was therefore provocative and raised the possibility that *Dh* may represent a mutant allele of *En1*.

Studies in this laboratory were undertaken by Higgins *et al.* (1992) to establish the genetic linkage distance between *Dh* and *En1*. A backcross analysis was performed segregating for *Dh* and polymorphic alleles of *En1* as well as three other closely linked markers, the endogenous murine leukaemia virus locus *Emv17*, the coat colour gene *leaden*, *ln*, and the red blood cell enzyme *Peptidase-3*, *Pep3*. Analysis of 563 offspring showed the most likely gene order and recombination frequencies for these markers to be *ln*(5.2±0.9) *Emv17*(1.1±0.4) *Dh*(0.7±0.4) *En1*(3.0±0.7) *Pep3*. These data show that *Dh* and *En1* are separable by recombination and cannot therefore be allelic. This result is in agreement with Martin *et al.* (1990) who, also by backcross analysis, demonstrated the same gene order of *Emv17* (0.28±0.28) *Dh* (0.28±0.28) *En1*. There is a discrepancy in the genetic distances reported between the two studies, but this is likely to result from the different inbred strains used in each study; Martin *et al.* generated heterozygous parental animals by mating *Dh* mice, maintained on a (C57BL/6J×C3HeB/FeJ)$F_1$ background, to RF/J mice, while Higgins *et al* generated heterozygous parental animals by mating *Dh* mice, maintained on an FZT background, to the inbred strain C57BL/6OlaWs. The proportion of male and female meiosies analysed was not significantly different between the two studies.

At the molecular level the data also supports the conclusion that *En1* is not an allele of *Dh*. *En1* transcripts are of normal size and abundance in *Dh* heterozygotes and homozygotes (Martin *et al.*, 1990). As well as being expressed in tissues affected by *Dh*, such as limb buds, ribs and vertebrae, *En1* is expressed in tissues which are apparently unaffected, such as the CNS (Davidson *et al.*, 1988; Davis and Joyner, 1988; Davis *et al.*, 1991). However no *En1* expression is present in the splanchnic mesoderm which is thought to be the primary site of action of the *Dh* mutation (Green, 1967). These studies however do not rule out the formal possibility that the *Dh* phenotype is caused by aberrant *En1* expression, perhaps through mutation of a regulatory element distantly located to *En1*.

### 3.1.3 Assignment of *Gli2* to the *Dh* region of chromosome 1

Recently a second gene of possible significance to *Dh* was found to be located within the *Dh* critical region. The chromosomal location of *Gli2*, a member of the *Gli* zinc finger family of genes, was initially assessed by fluorescent in situ hybridisation (FISH) to metaphase spreads of mouse chromosomes and found to lie in band 1E in the central region of chromosome 1 (Hughes *et al.*, 1997). Work in this laboratory by David Hughes (Hughes *et al.*, 1997) subsequently determined a more precise location of *Gli2*. A microsatellite marker was isolated from a genomic clone containing an intron within the *Gli2* gene. This marker proved highly polymorphic enabling it to be mapped on the BXD and the BXH recombinant inbred strain, RI, panels. The strain distribution pattern obtained indicates *Gli2* is located between *Emv17* and *En1*, within the *Dh* region, illustrated in figure 3.1. The position of *Gli2* relative to *Dh* was then determined by analysis of animals from the *Dh* backcross which were recombinant between *Emv17* and *En1*. As shown in figure 3.3, *Gli2* is non-recombinant with *Dh*, identifying *Gli2* as a candidate gene for *Dh* (the candidacy of *Gli2* for *Dh* is fully discussed in chapter 4). Subsequently a sequence variation was detected in the 3′ untranslated sequence of *Gli2*; a run of 8 Gs was observed in the *Dh* allele compared to 12 Gs in C57BL/6 (the inbred strain used in the generation of heterozygous *Dh* mice used in the *Dh* backcross analysis). When the relevant backcross progeny were typed for this marker recombination with *Dh* was observed in three of the progeny (figure 3.3). The 3′ of *Gli2* therefore delimits the proximal end of the region in which *Dh* could lie, *En1* delimiting the distal end. The recombination fractions observed are: *Gli2* 3′- *Dh* 3/562 = 0.5±0.3cM, *Dh-En1* 4/563 = 0.7±0.4cM. The overall *Dh* critical region is therefore 1.2±0.5cM.

(Two animals from the *Dh* backcross, one recombinant between *Emv17* and *Dh* and one recombinant between *Dh* and *En1*, are no longer available for analysis. The recombination fraction stated above, between *Gli2* 3′and *Dh*, is therefore 3/562 and not 3/563).

### 3.1.4 Aims: Cloning and physical characterisation of the *Dh* critical region

The delineation of *Dh* to a defined genetic interval of only 1.2±0.5cM provides the information to enable the physical cloning of the region containing the *Dh* mutation. The genetic map of the mouse is estimated at 1600cM while the haploid genome size is approximately 3 x $10^9$ bp. Assuming a constant relationship between genetic and physical distance 1cM would therefore equal approximately 2Mb. The *Dh* region would hence encompass approximately 2.4Mb. However it is observed that in different regions of the genome this simple linear relationship is not adhered to in a regular manner. For example Barsh and Epstein (1990) report values of <150kb/cM and 4Mb/cM in two adjacent segments of the *agouti* locus. Based on this data the physical extent of the *Dh* region could encompass between 180Kb to 4.8Mb.

The aim of the work presented in this chapter was to clone the genomic region containing *Dh* by the isolation of a series of overlapping YAC clones, a YAC contig. This would provide the resource to physically characterise the region and identify candidate genes for *Dh*. Initially *Gli2* and *En1* markers were used to isolate YAC clones at each end of the contig. Markers subsequently derived from these YACs were then used to isolate or 'walk' to YAC clones internal to the region. This procedure can be repeated until YACs from either end of the contig are shown to overlap. In addition to the YAC based map, mouse genomic DNA was analysed by pulsed field gel electrophoresis (PFGE), to generate a long range genomic restriction map of the *Dh* region. This would aid in providing physical linkage between markers and confirm the accuracy of the YAC contig. A further aim was to generate markers which would allow integration of the emerging physical map with the genetic map. This would provide corroboration for the accuracy of each map and may enable the size of the *Dh* interval to be reduced. In addition relevant genes and anonymous markers arising in the literature were investigated for their possible association with the *Dh* region.

## 3.2 Results

### 3.2.1 Genetic localisation of *Inhibinβb* to the *Dh* region

*Inhibinβb*, *Inhβb*, a member of the *TGFβ* superfamily, has been localised to chromosome 1 by Lafuse and Zwilling (1993). They analysed the distribution of *Inhβ b* in the BXD RI panel and found full concordance between *Inhβb* and *En1*. *Gli2*, also placed on the BXD panel, is found to lie proximal to *En1* and *Inhβb* but distal to *Emv17*, as shown in figure 3.1 (Hughes *et al.*, 1997). This localisation of *Inhβb* was intriguing because *Inhibinβa, Inhβa,* maps to chromosome 13 in the vicinity of *Gli3* (Barton *et al.*, 1989). This indicates that *Gli2-Inhβb* and *Gli3- Inhβa* represent paralogous linkage groups. This region of paralogy does not however include paralogous *engrailed* genes as there is no *engrailed* gene on chromosome 13. Hence because *Inhβb* co-segregates with *En1* and not *Gli2* it is likely that *Inhβb* is located on the *Gli2* side of *En1*. This would place *Inhβb* within the *Dh* region.

To investigate the position of *Inhβb* relative to the *Dh* region a polymorphic marker was generated such that the informative recombinant mice produced from the *Dh* backcross could be typed. An *Inhβb* specific probe was generated by PCR amplification of a 671bp fragment from C57BL/6 genomic DNA using primers I181/I182 (designed from mouse *Inhβb* cDNA sequence, Ritvous *et al.*, 1995). Sequence analysis confirmed this fragment to be *Inhβb*. This probe was used to screen a 129/Sv mouse genomic lambda library (Stratagene) resulting in the isolation of one clone. In an attempt to isolate microsatellite repeats which may provide a polymorphic marker this clone was digested with *Sau3A*I and shotgun cloned into pBluescript. Approximately 500 colonies (a three fold redundancy) were screened with a $(GT)_{11}$ oligonucleotide, using the method of Buluwela *et al.* (1989). Stringent washing conditions were used in order to identify uninterrupted CA repeats which are more likely to be polymorphic. Two positive colonies were isolated and the

**Figure 3.1** Strain distribution patterns of loci in the *Dh* region in the BXD recombinant inbred strains. Markers placed previous to this study are; *Emv17* (Buchberg *et al.*, 1986), *En1* (Hill *et al.*, 1987), *Inhβb* (Lafuse and Zwilling, 1993), *D1Ncvs44* and *D1Ncvs45* (Hayashizaki *et al.*, 1994), *Gli2*, *D1Mit417*, *D1Mit258*, *D1Mit285*, *D1Mit188*, *D1Mit389*, *D1Mit419*, *D1Mit217* and *D1Mit88* (Hughes *et al.*, 1996). The SDP for the YAC end clone *127B7RE* is illustrated, the marker being non-concordant with the *Dh* region of chromosome 1.

| Marker | Animal | | | | | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 5 | 6 | 8 | 9 | 11 | 12 | 13 | 14 | 15 | 16 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 27 | 28 | 29 | 30 | 31 | 32 |
| *Emv17* | B | B | D | D | B | B | D | B | D | B | D | D | D | B | D | B | D | D | D | B | B | D | D | D | B | D |
| *D1Mit417* | B | B | D | D | B | B | D | B | D | B | D | D | D | B | D | B | D | D | D | B | B | D | D | D | B | D |
| *D1Mit258* | B | B | D | D | B | B | D | B | D | B | D | D | D | B | D | B | D | D | D | B | B | D | D | D | B | D |
| *D1Mit285* | B | B | D | D | B | B | D | B | D | B | D | D | D | B | D | B | D | D | D | B | B | D | D | D | B | D |
| *Gli2* | B | B | D | D | B | B | D | B | D | B | D | D | D | B | D | B | X<br>B | D | D | B | B | D | D | D | B | D |
| *67G3LE* | B | B | D | D | B | B | D | B | D | B | D | D | D | B | D | B | B | D | D | B | B | D | D | D | B | D |
| *D1Mit188* | B | B | D | D | B | B | D | B | D | B | D | D | D | B | D | B | B | D | D | B | B | U | D | D | B | D |
| *D1Ncvs44* | B | B | D | D | B | B | D | B | D | B | D | D | D | B | D | B | B | D | D | B | B | D | D | D | B | D |
| *D1Ncvs45* | B | B | D | D | B | B | D | B | D | U | D | D | D | B | D | B | B | D | D | B | U | U | D | D | B | D |
| *Inhβb* | B | B | D | D | B | B | D | B | D | B | D | D | D | B | D | B | B | D | D | B | X<br>D | D | D | D | B | D |
| *En1* | B | B | D | D | B | B | D | B | D | B | D | D | D | B | D | B | B | D | D | B | D | D | D | D | B | D |
| *D1Mit88* | B | B | D | D | B | B | D | B | D | B | D | D | D | B | D | B | B | D | D | B | D | D | D | D | B | D |
| *D1Mit389* | B | B | D | D | B | B | D | B | D | B | D | D | D | B | D | B | B | D | D | B | D | D | D | D | B | D |
| *D1Mit217* | B | B | D | D | B | B | D | B | D | B | D | D | D | B | D | B | B | D | D | B | D | D | D | D | B | D |
| *D1Mit419* | B | B | D | D | B | B | D | B | D | B | D | D | D | B | D | B | B | D | D | B | D | D | D | D | B | D |
| *129B7RE* | B | D | B | B | D | B | D | U | D | B | D | B | B | B | U | U | D | U | B | D | B | D | U | B | U | B |

plasmid inserts sequenced. In both clones the same uninterrupted $(CA)_{18}$ repeat was identified indicating the presence of just one CA repeat in the lambda clone. PCR primers I734 and I735, which amplify across the repeat, were designed. Amplification from *Dh* homozygotes and C57BL/6 identified a size polymorphism, fragments of 157bp and 141bp being produced respectively.

Eight backcross mice recombinant in the *Dh* region, between *Emv17* and *En1*, were available for analysis (see chapter 2, 2.1). These were typed for the *Inhβb* assay as shown in figure 3.2. Comparison of the *Inhβb* genotype for each animal with the genotype of *Dh* flanking markers shows *Inhβb* to be genetically inseparable from *Dh*, as illustrated in figure 3.3. This analysis, in agreement with the above reasoning, confirms *Inhβb* to lie proximal to *En1*, due to recombination demonstrated in three of the backcross progeny, while the separation of *Inhβb* and *Gli2* on the BXD panel locates *Inhβb* distal to *Gli2*. *Inhβb* is therefore located within the genetically defined region containing the *Dh* gene and as such represents a candidate gene for *Dh*. The analysis of *Inhβb* in relation to the *Dh* mutation is presented in chapter 4. Further *Inhβb* represents a valuable marker which has been exploited in the construction of the YAC contig and genomic map of the *Dh* region, as described in sections 3.2.7 and 3.2.8 below.

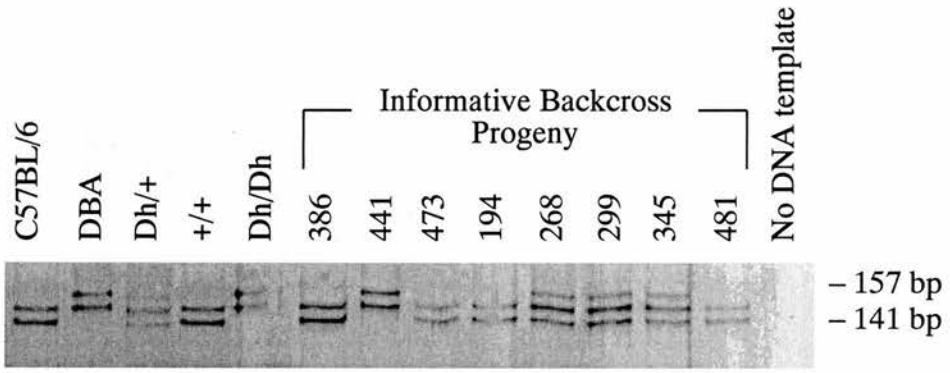
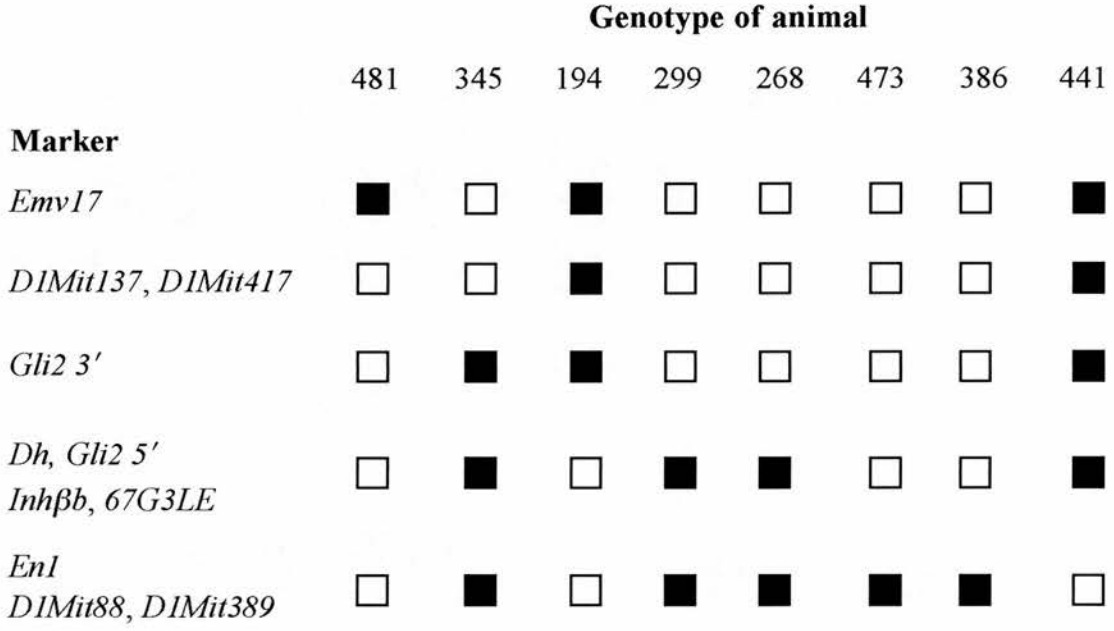



**Figure 3.2 Genetic localisation of *Inhβb* relative to *Dh***

Silver stained 6% polyacrylamide gel showing the genotype of the informative *Dh* backcross progeny for the *Inhβb* microsatelite assay I734/I735. The *Dh* chromosome produces a band of 157bp compared to a 141bp C57BL/6 band. A background band of approximately 150bp is also seen in each lane, however as this band is amplified from each strain it does not affect the analysis.

**Figure 3.3** Segregation of markers in the informative progeny from the *Dh* backcross recombinant between *Emv17* and *En1*. Each column represents the chromosome that was inherited from the (*Dh*/+ x C57BL/6)F1 parent. Black boxes represent the *Dh* allele and white boxes represent the C57BL/6 allele.




### 3.2.2 Placement of MIT microsatelite markers relative to *Dh*

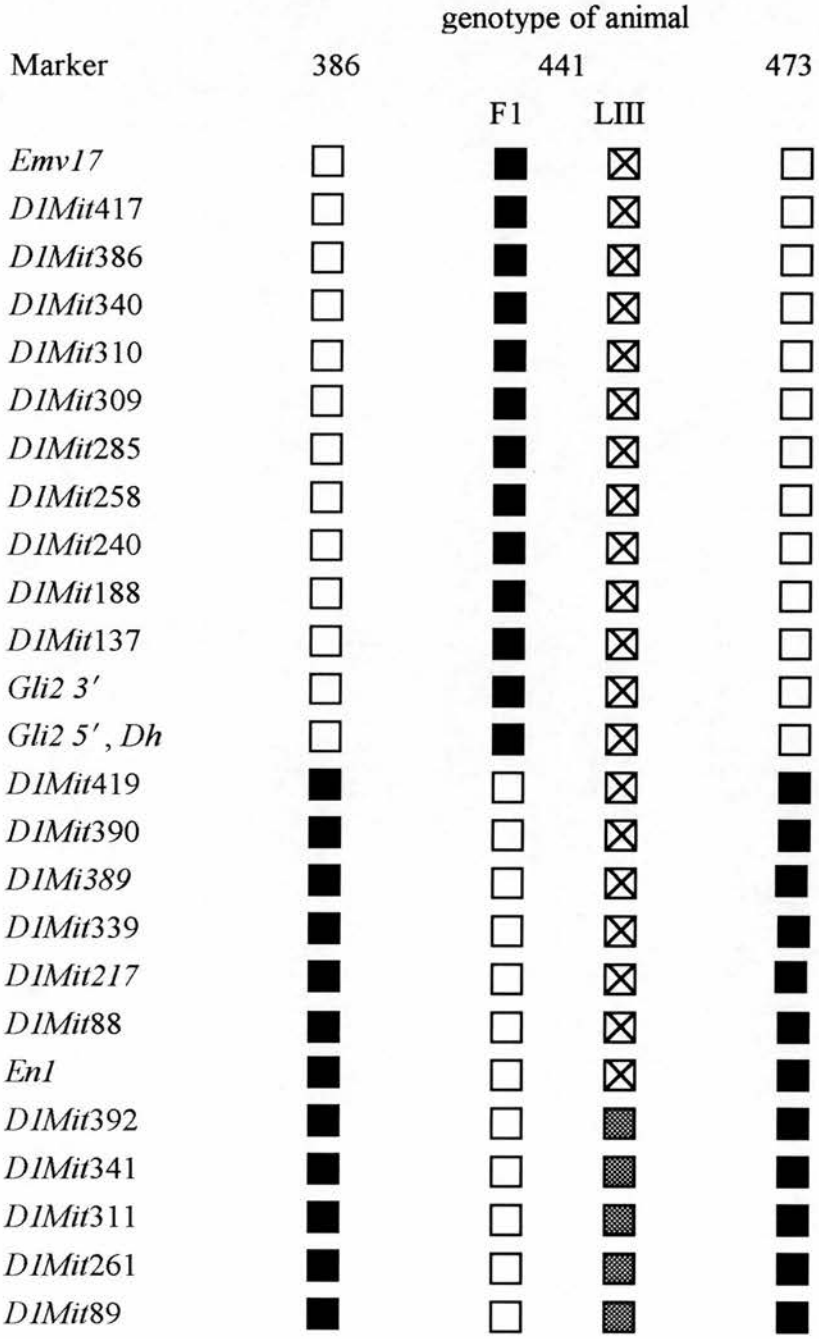
The closest flanking *D1Mit* markers to *Dh* had been determined to be *D1Mit417*, *D1Mit137* proximally and *D1Mit88*, *D1Mit389* distally (Hughes *et al.*, 1997). At the time of this study 35 *D1Mit* markers were available which had been placed in-between or at the same genetic location as *D1Mit417* and *D1Mit88* (MIT release 6.2, 1994). Potentially therefore one or more of these markers could map to the *Dh* region and prove informative. Of these markers, 27 were polymorphic between C57BL/6 and either DBA or C3H (*Dh* arose on either a C3H, CBA or 101 background but the *Dh* chromosome is identical to DBA and C3H at all loci tested) and could thus potentially be placed on the panel of informative recombinants from the *Dh* backcross.

The 27 markers (purchased from Research Genetics) were assayed by PCR for polymorphism between the C57BL/6 and *Dh* alleles. 21 were determined to give a

reliable polymorphism which could be used to type the recombinant mice. Initially only those mice recombinant between *Dh* and *En1*, (progeny 386, 441, and 473) were typed to identify those markers which lay distal to, on the *En1* side of, the recombination breakpoints. Those which are shown to lie proximal, on the *Gli2* side, of the recombinations could then be typed in the mice recombinant between *Dh* and *Emv17*. The results are shown in figure 3.4.

All of the markers type either proximally or distally of the recombinations in all three animals. Therefore no markers are located in the region between the three recombination break points. Eleven markers, 88, 89, 217, 261, 311, 339, 341, 389, 390, 392 and 419 are located on the *En1* side of the three recombination break points. Five of these, 89, 261, 311, 341 and 392 can be located distally to *En1* due to the genotype observed in the LIII chromosome (the outbred strain to which (*Dh*/+ x C57BL/6)F1 mice were crossed in the *Dh* backcross) of animal 441. Markers 88, 217, 339, 389, 390 and 419 are non-recombinant with *En1* and may therefore lie on either side of *En1*. *D1Mit*217is however derived from the *En1* locus itself. Should any of these lie on the *Dh* side of *En1* they would limit the physical region in which the *Dh* gene can reside. To address this point YACs isolated from the Dh region (see 3.2.3 below) were assayed by PCR for their presence. As shown in figure 3.5 *D1Mit*390 is present in the *En1* containing YAC 130G11 only while *D1Mit*419 is present in all three *En1* containing YACs 114C12, 130G11 and 175D3. Physical location of these two markers on the relevant YACs was attempted by probing Southern blots of restricted YAC DNA with the *D1Mit* oligonucleotides which had been $^{32}P$ end labelled. Localisation of D1Mit419 was unsuccessful but *D1Mit*390 was demonstrated to lie within a 25kb region 100kb from the left arm of 130G11 as shown in figure 3.17. As the transcriptional orientation of *En1* relative to the contig is not known *D1Mit*390 may lie either distal or proximal to *En1*. If it is shown to lie proximal, *D1Mit*390 would define a new minimal region in which the *Dh* mutation can lie.

**Figure 3.4** Segregation of *D1Mit* markers from the *Dh* region in the three *Dh* backcross progeny recombinant between *Dh* and *En1*. Each column represents the chromosome that was inherited from the (*Dh*/+ x C57BL/6)F1 parent. Black boxes represent the *Dh* allele and white boxes represent the C57BL/6 allele. For animal 441 the Genotype of the LIII chromosome is also illustrated to demonstrate markers located distal to *En1*. Crossed boxes represent a DBA like allele and shaded boxes represent a C57BL/6 like allele.



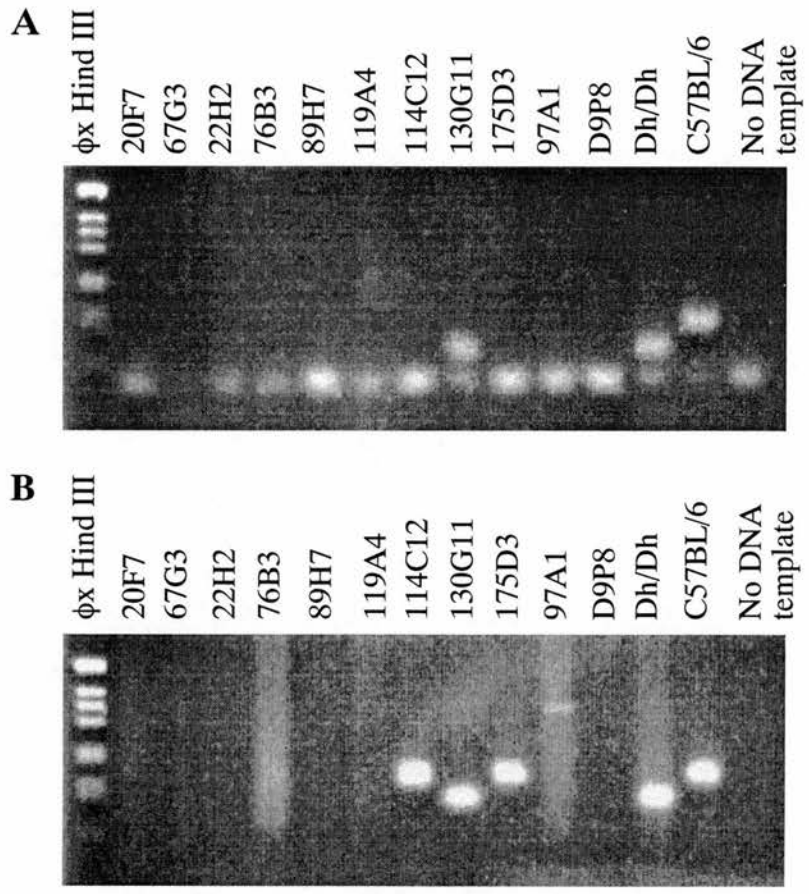



**Figure 3.5 Identification of YACs containing *D1Mit* markers**

**A**. 3% agarose gel showing amplification products of the *D1Mit390* marker amplified from *Dh* region YACs. *D1Mit390* only amplifies from YAC 130G11.

**B**. 3% agarose gel showing amplification products of the *D1Mit419* marker amplified from *Dh* region YACs. *D1Mit419* amplifies from YACs 114C12, 130G11 and 175D3.

The difference in size of the amplified bands, demonstrates the polymorphic nature of the marker; YACs 114C12 and 175D3 are C57BL/6 derived while 130G11 is *C3H* derived.

The markers which type on the *Gli2* side of the recombinations, 137, 188, 309, 310, 240, 258, 285, 340 and 368 could either lie distally to *Gli2*, within the *Dh* region or proximally to *Gli2*. Because three mice are recombinant within *Gli2*, mice 194, 268 and 299, the relative position of these markers to *Gli2* can be determined by typing these mice. Markers which lie on the *Dh* side of a recombination in these mice will be located distally to the 3′ of *Gli2*. Although these markers cannot improve the genetic resolution of the map surrounding *Dh*, because *Gli2* is non-recombinant with *Dh*, they may be useful in constructing the physical map of the region. This analysis has not been performed.

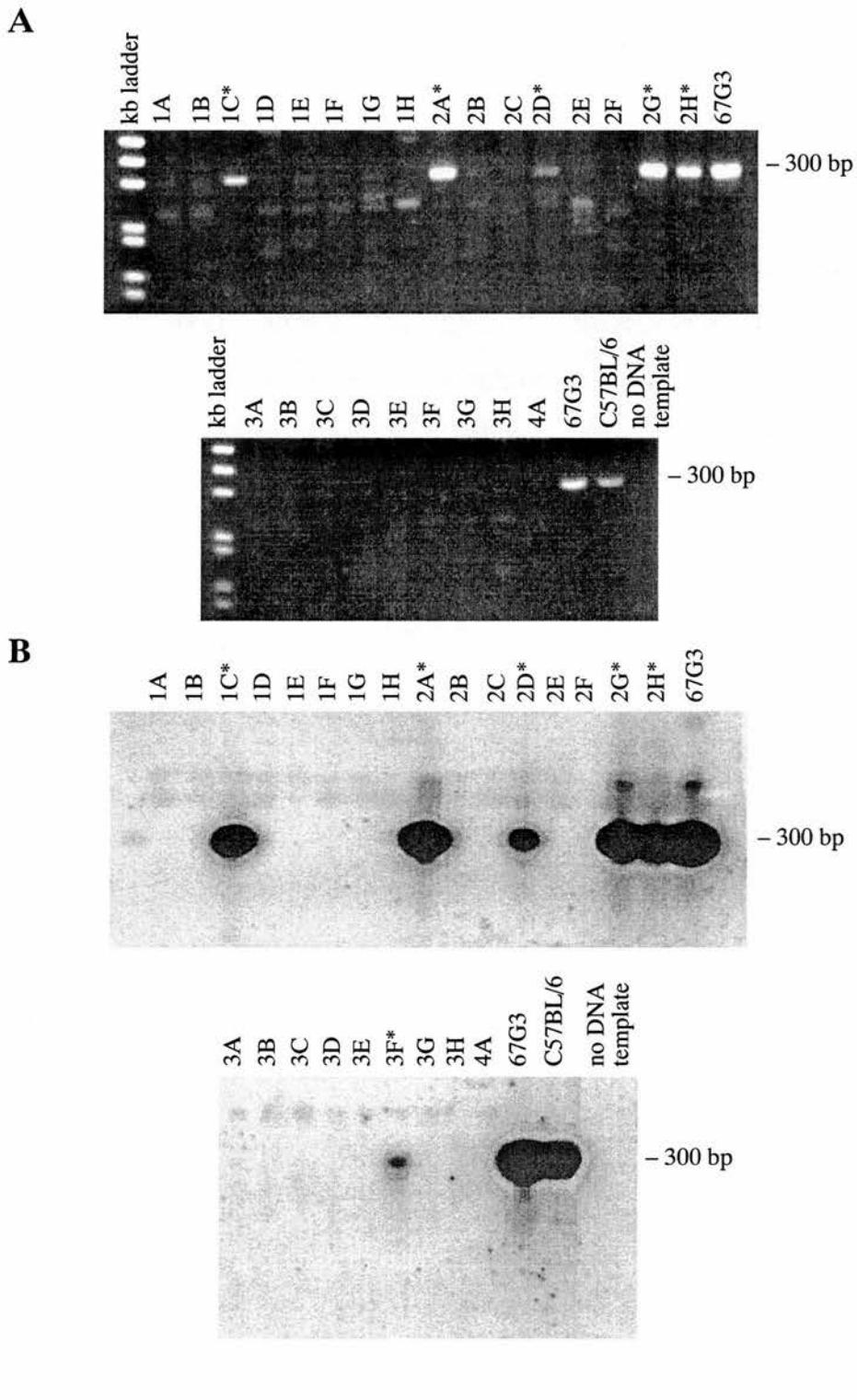
### 3.2.3 Isolation of YACs from the *Dh* region

Three YAC libraries were available at the time of this study. The St. Marys' Hospital library (Chartier *et al.*, 1992) and the ICRF library (Larin *et al.*, 1991) were available in a combined format for screening by PCR. The Whitehead/MIT library (Kusumi *et al.*, 1993) was also available for screening by PCR. Library details and screening protocols are described in chapter 2 (2.6.3).

The three libraries were initially screened to isolate clones containing the two closest flanking markers to *Dh*, *Gli2* (PCR primers 511/512) and *En1* (PCR primers 257/829). Two YACs were isolated; 129B7, screening with 511/512 and 114C12, screening with 257/829. Given the three libraries combined represent approximately eleven genome equivalents, the isolation of only one clone for each marker is a poor return. Therefore to attempt the isolation of more clones for these markers additional screens were performed using alternative PCR primers; *Gli2*: I51/1C, I34/I35 and F264/F265, *En1*: 990/989. Only one additional clone was isolated; 67G3 screening with I51/1C. Analysis of these YACs showed that they did not span the *Dh* region (see 3.2.7 below). End clone markers were therefore derived from these YACs to enable rescreening of the libraries (see 3.2.4 below). This enables the identification of YACs which overlap those previously isolated, enabling the contig to be extended.

The results of screening with the *Gli2* and *En1* markers and all subsequently derived end clone markers are shown in table 3.1. A total of twelve YACs were isolated.

As an example of the screening procedure, figures 3.6 and 3.7 illustrate typical results of screening the primary and secondary pools respectively of the Whitehead/MIT library for the end clone derived marker 67G3LE. Figure 3.5 shows the primary pool screen and the identification of five positive secondary pools, while figure 3.6 shows the identification of the three co-ordinates from secondary pool 12 which identify an individual clone, in this case clone 89H7. Frequently false positive and negative results were obtained. For example primary pools were identified which failed to give any secondary co-ordinates. This could result from contamination of the primary pool DNA giving a false positive result. If however the identification of the primary pool represents a genuine positive then failure to locate secondary co-ordinates is likely to be due to a failure of the positive clone to be sufficiently represented in the secondary pool DNAs or to degradation of the DNA. Screening of the secondary pools often identified multiple plate, row or column co-ordinates. Again this is likely to be due to contamination of the secondary pool DNAs. When this occurred all possible positive clones were individually screened. Twice, however multiple positive secondary pool co-ordinates were obtained, for markers *Gli2* I51/1C and 67G3LE 580/581, but which failed to yield a positive clone. This suggests that the identification of the primary pool represents a genuine clone but that combined false positive and false negative results failed to identify the correct secondary co-ordinates. Alternatively the correct co-ordinates may have been obtained, the individual clone escaping identification due to the deletion of the marker from the YAC during propagation. On many occasions secondary pool screening failed to yield any positive signal for one or more co-ordinates. When only one co-ordinate was missing it is possible to screen all potential positive clones, for example YAC 119A4 was identified in this way. However this is not feasible if two co-ordinates are missing.


A
kb ladder
1A
1B
1C*
1D
1E
1F
1G
1H
2A*
2B
2C
2D*
2E
2F
2G*
2H*
67G3
– 300 bp
kb ladder
3A
3B
3C
3D
3E
3F
3G
3H
4A
67G3
C57BL/6
no DNA template
– 300 bp
B
1A
1B
1C*
1D
1E
1F
1G
1H
2A*
2B
2C
2D*
2E
2F
2G*
2H*
67G3
– 300 bp
3A
3B
3C
3D
3E
3F*
3G
3H
4A
67G3
C57BL/6
no DNA template
– 300 bp

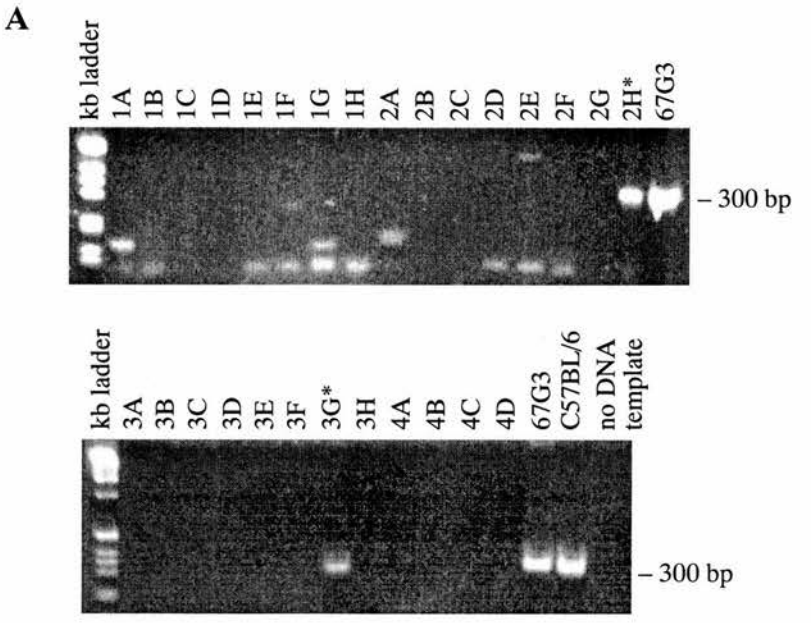

A
kb ladder
1A
1B
1C
1D
1E
1F
1G
1H
2A
2B
2C
2D
2E
2F
2G
2H*
67G3
– 300 bp
kb ladder
3A
3B
3C
3D
3E
3F
3G*
3H
4A
4B
4C
4D
67G3
C57BL/6
no DNA template
– 300 bp
B
1A*
1B
1C
1D
1E
1F
1G
1H
2A
2B
2C
2D
2E
2F
2G
2H*
67G3
– 300 bp
3A
3B
3C
3D
3E
3F
3G*
3H
4A
4B
4C
4D
67G3
C57BL/6
no DNA template
– 300 bp

**Table 3.1 YAC library screening results for markers in the *Dh* region**

For each marker all primary pools which gave a positive signal are indicated (for end clone markers the co-ordinates of the YAC from which the end clone was derived are not however repeated). End clone markers are indicated by the YAC name and LE, or RE, for left and right end clone respectively. All secondary pool co-ordinates obtained are shown in the order plate or floor, row, column. Multiple co-ordinates identified for any axis are shown and failure to identify a co-ordinate is indicated by -. All successfully isolated YAC clones are listed. * The 175D3LE assay, 261/262, identified clone 99C8 from the Whitehead/MIT library, however subsequent analysis failed to yield a positive clone for further analysis.

| Marker / PCR primers | ICRF / St. Marys' Library | | | | | Whitehead/MIT library | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | 1° pool identified | 2° pool co-ordinates | | | YAC isolated | 1° pool identified | 2° pool co-ordinates | | | YAC isolated |
| | | plate | row | column | | | plate | row | column | |
| *En1*, 257/829 | - | - | - | - | - | 11 | 8 | A | 1 | 114C12 |
| | | | | | | 22 | - | - | - | - |
| 114C12LE | 8 | κ, ζ | J, O | 1, 19, | 130G11 | 9 | 8 | - | - | - |
| 358/360 | | | | 23 | | 13 | C | - | - | - |
| | | | | | | 15 | D | - | - | - |
| | | | | | | 22 | 7, | D | 3 | 175D3 |
| 175D3LE | 8 | κ, ζ | A, O | 1, 23 | 97A1 | 8 | 2 | B | - | - |
| 261/262 | 4 | - | - | - | - | 9 | 8 | C | 4 | 72C4 |
| | | | | | | 13 | 3 | C, H | 8 | (99C8)* |

**Table 3.1 YAC library screening results for markers in the *Dh* region (cont.)**

For each marker all primary pools which gave a positive signal are indicated (for end clone markers the co-ordinates of the YAC from which the end clone was derived are not however repeated). End clone markers are indicated by the YAC name and LE, or RE, for left and right end clone respectively. All secondary pool co-ordinates obtained are shown in the order plate or floor, row, column. Multiple co-ordinates identified for any axis are shown and failure to identify a co-ordinate is indicated by -. All successfully isolated YAC clones are listed.

| Marker / PCR primers | ICRF / St. Marys' Library | | | | | Whitehead/MIT library | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | 1° pool identified | 2° pool co-ordinates | | | YAC isolated | 1° pool identified | 2° pool co-ordinates | | | YAC isolated |
| | | plate | row | column | | | plate | row | column | |
| *Gli2*, 511/512 | 7 | ζ | H | 2 | 129B7 | - | - | - | - | - |
| *Gli2*, I51/1C | 7 | ζ | H | 2 | 129B7 | 3 | - | - | - | - |
| | | | | | | 8 | 2, 4, 6, 8 | D, F | 2, 4 | - |
| | | | | | | 9 | 3, 8 | G | 3, 4, 5 | 67G3 |
| | | | | | | 16 | - | - | - | - |
| | | | | | | 20 | - | - | 9 | - |
| 67G3LE 580/581 | 8 | α | J | 15 | 76B3 | 3 | 4 | D, F | 7 | 20F7 |
| | | | | | | 12 | 1 | H | 7 | 89H7 |
| | | | | | | 15 | 7 | - | 4 | 119A4 |
| | | | | | | 16 | 8 | B | 2, 5, 6 | - |
| | | | | | | 22 | - | - | - | - |

### 3.2.4 Isolation of YAC end clones

Isolation of a YAC end clone, the terminal sequence of the recombinant DNA cloned within a YAC, provides a versatile marker. End clones can be used to rescreen YAC libraries for the purposes of building a YAC contig and to assess marker content of potentially overlapping YACs. Generation of an informative end clone polymorphism enables genetic localisation, allowing assessment of YAC chimerism and integration of the physical and genetic maps.

Numerous methods have been described for end clone isolation. These include strategies based on the generation of individual YAC libraries via subcloning into cosmid, lambda or plasmid vectors. End clones are then selected by hybridisation with the YAC vector arm probes. For the technique of 'junction trapping' (Patel *et al*, 1993) YAC is subcloned into plasmids and the end clones generated by PCR amplification using a combination of plasmid and YAC vector primers. These techniques however require a cloning step which can be laborious and time consuming. For this reason PCR based techniques are more commonly used. All the PCR based techniques make use of primers derived from the YAC vector arm to provide the end-specificity. The other primer may be derived from a linker such as a vectorette (Riley *et al*., 1990) or splinkerette (Devon *et al*., 1995), a consensus repeat sequence such as B1 or B2 (Herman *et al*., 1992) or for the inverse PCR method, another primer from the YAC vector (Ochman *et al*, 1988, Arveiler and Porteous, 1991).

The speed and specificity of the inverse PCR protocol described by Arveiler and Porteous, (1991) made it the method of choice. The method is described in chapter 2 (2.6.7). Examples of this technique are shown in figures 3.8 and 3.9 for the isolation of right and left arm end clones respectively. Table 3.2 lists end clones successfully isolated from the YACs in the *Dh* region. The failure to isolate some end clones is likely due to inconveniently located restriction sites in the YAC inserts which cause failure of the PCR under the conditions used. In these cases the vectorette method of Riley *et al* (1990) was adopted but in all cases this also failed.

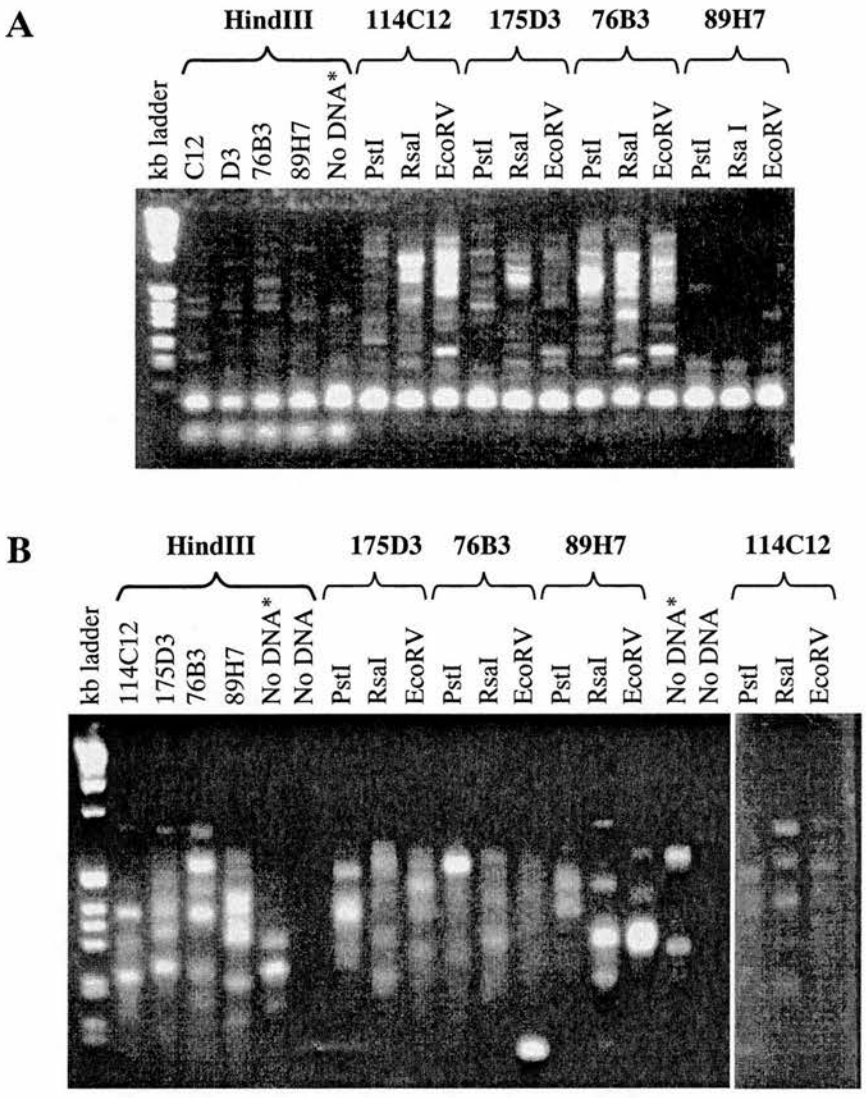



**Figure 3.8 Amplification of YAC right-hand end clones by inverse PCR**

**A** Inverse PCR amplification of the right hand end clones from YACs 114C12, 175D3, 76B3 and 89H7. PCR primers 375/556 were used in conjunction with *Hind*III while primers 374/556 were used for *Pst*I, *Rsa*I and *EcoR*V digested DNA.

**B** Nested PCR of the reactions shown in panel A. PCR primers C627/375 for *Hind*III and primers 626/374 for *Pst*I, *Rsa*I and *EcoR*V digested YAC DNA. (The lane labelled No DNA* indicates nested amplification from the No DNA* lane in panel A). From this experiment the bands of 114C12 *Rsa*I 800bp and 76B3 *Pst*I 600bp were taken for further analysis and represented genuine end clones.

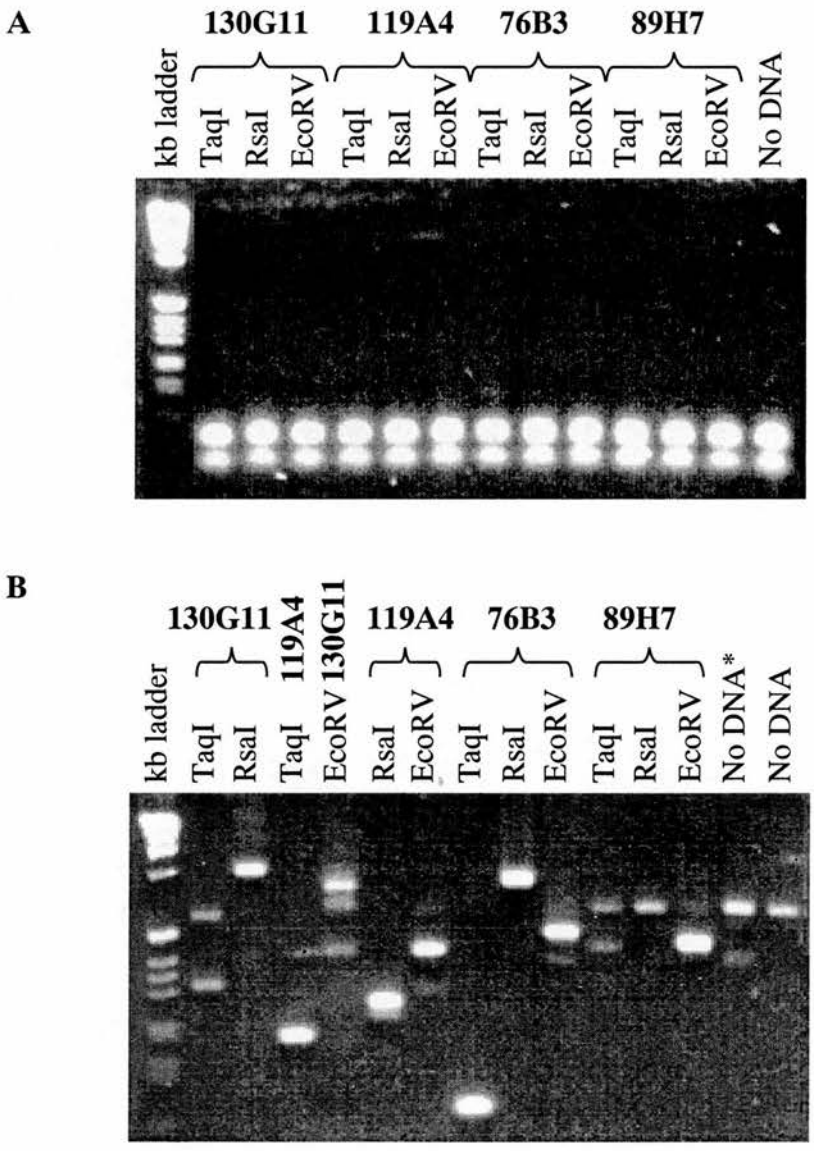



**Figure 3.9 Amplification of YAC left-hand end clones by inverse PCR.**

**A** Inverse PCR Amplification of the left hand end clones from YACs 130G11, 119A4, 76B3 and 89H7 using primers 372/373.

**B** Nested PCR of the reactions showed in panel A using primers C626/J105.

(The lane labelled No DNA* indicates nested amplification from the No DNA lane in panel A).

From this experiment the bands of 130G11 *Rsa*I 1kb and 76B3 *Rsa*I 950bp were taken for further analysis and represented genuine end clones.

To generate versatile markers from these end clones, sequence was generated from each fragment such that PCR primers could be designed which would amplify from genomic DNA. The use of nested PCR to generate an end clone and the ability to read through YAC vector sequence into genomic sequence confirmed the origin of the amplified products as being the terminal sequences from YAC inserts. PCR amplification from YAC template DNA using the end clone PCR primers also confirmed their origin.

**Table 3.2 End clones isolated from YACs**

| YAC (LE, left hand end RE, right hand end) | | Size of end clone | PCR primers designed |
|---|---|---|---|
| 129B7 | LE | - | - |
| | RE | 1.5kb | H58/H59 |
| 67G3 | LE | 320bp | I580/I581 |
| | RE | - | - |
| 119A4 | LE | - | - |
| | RE | 340bp | I975/I976 |
| 89H7 | LE | - | - |
| | RE | 300bp | K680/K682 |
| 76B3 | LE | 950bp | J141/J142 |
| | RE | 600bp | J145/J146 |
| 114C12 | LE | 250bp | H358/H360 |
| | RE | 800bp | J114/J115 |
| 130G11 | LE | 1.0kb | J143/J144 |
| | RE | 1.2kb | I143/I144 |
| 175D3 | LE | 580bp | I261/I262 |
| | RE | - | - |

### 3.2.5 Assessment of YAC chimerism

An important characteristic to determine of any YAC to be analysed is that the cloned DNA is a contiguous fragment which truly represents the genome. Chimeric YACs, containing DNA from disparate regions of the genome, are however present at rates of 40% to 50% in all the available libraries. Two main mechanisms are thought to be responsible for chimerism which occur during the construction of the library. Firstly two or more genomic fragments may co-ligate during the ligation of vector arms to the genomic fragments. Secondly co-transformation of two or more YACs into the same yeast cell can result in recombination generating chimeric clones (Green *et al.*, 1991).

Chimerism can be assessed in a number of ways. YAC DNA may be hybridised to metaphase chromosome spreads via fluorescent in situ hybridisation, (FISH) technology (Senger, 1995). Any signal other than from the expected chromosomal region indicates chimerism. If a number of YACs have been isolated from the same region, the contiguous YACs will have the same restriction map and marker content, whereas chimeric YACs should be detected by a divergent restriction pattern or an absence of appropriate markers. Similarly if a physical genomic map is available for a region then the physical map of a YAC should concur. Detection of suspect YACs using restriction mapping and marker content may however mean rearrangement and/or deletion of regions in the YAC, rather than chimerism. Isolation of DNA fragments from the ends of YAC inserts can be used to type somatic cell hybrid panels which will indicate the chromosomal origin of a marker (Williamson *et al.*, 1995). These methods are all valuable but can possibly give false results. For example restriction mapping and FISH analysis may not detect small chimeric regions (see chapter 5, 5.3.2) and somatic cell hybrid panel analysis will usually only indicate the chromosome of origin for a marker and may not therefore resolve intra-chromosomal chimerism. The most definitive way to detect chimerism is to genetically map markers derived from the ends of YACs. This provides precise

localisation for the very terminal sequence of the YAC, chimerism being revealed by deviation from the expected map position.

In this study YAC chimerism was assessed using a combination of the above. Firstly detailed restriction maps were generated for each YAC allowing YACs to be aligned relative to one another (see 3.2.7 and figure 3.18). Comparison of restriction sites and marker content enabled chimerism to be positively detected in three YACs. The restriction sites in the right hand 200kb of YAC 129B7 do not correspond to either those in YACs 119A4 or 76B3 or to the sites placed on the genomic map. Similarly 75kb at the right hand end of 20F7 does not contain those sites present on the genomic map or in YACs 119A4, 76B3 and 89H7. Hybridisation and PCR analysis (data Not shown) also demonstrate the absence of inhibinβb from 20F7. Although restriction sites could not be accurately mapped to the right end of 175D3 absence of D1Mit390 indicates chimerism.

The YAC 114C12 was analysed for chimerism by FISH. The other YACs were not analysed as they were isolated at a date subsequent to the FISH analysis. YAC DNA was prepared by 'catch-linker PCR' (Shibasaki *et al.*, 1995), as detailed in section 2.9.3., and used as probe by Muriel Lee for FISH analysis according to Fantes *et al* (1995). Figure 3.10 shows representative results observed in 12 metaphase spreads. The 114C12 probe produces a unique signal which is located in the central region of chromosome one. This is in concordance with the genetic location of *En1* (Joyner and Martin, 1987; Hill *et al.*, 1987; Martin *et al.*, 1990; Higgins *et al.*, 1992). As no other signal is seen at any other chromosomal location 114C12 was reasoned to be a non-chimeric YAC.

Detection of chimerism was further attempted by the identification of polymorphisms within end clones to enable the genetic position of the marker to be determined through the analysis of suitable genetic crosses. The BXD panel was chosen for the mapping of end clones because a strain distribution pattern was known for markers in the *Dh* region. Because the *Dh* chromosome is most similar to

**Figure 3.10 Fluorescent in situ hybridisation of YAC 114C12 to metaphase chromosome spreads**

114C12 probe was prepared by catch linker PCR and labelled with biotin. The biotin signal was detected with FITC against propidium iodide counter stained chromosomes. The FITC signal, arrowed, is located uniquely to the central region of chromosome one.

DBA, polymorphisms detected between C57BL/6 and DBA could also be mapped relative to *Dh* on the *Dh* backcross.

To identify polymorphisms, fragments were amplified from C57BL/6 and DBA with end clone specific primers. The products were then digested with a series of frequent cutting enzymes to identify restriction sites polymorphic between the two strains. Initially ten enzymes were used: *Sau*3AI, *Alu*I, *Hae*III, *Hha*I, *Hpa*II, *Rsa*I, *Sty*I, *Mae*II, *Taq*I and *Msp*I. Additional enzymes could be assayed if the above proved non-informative.

Using this technique polymorphisms were detected in two end clones. *Alu*I digestion of 129B7RE amplified from DBA gave fragments of 500bp, 250bp, 250bp, 210bp, 190bp, 160bp and 75bp whereas the product amplified from C57BL/6 gave fragments of 500bp, 475bp, 250bp, 220bp and 190bp. The strain distribution pattern (SDP) in the BXD panel was determined as shown in figure 3.11, the polymorphism being scored by the presence or absence of bands at 475bp and 220bp. This pattern is listed in figure 3.2 and is 55% concordant (11/20) with that for *Gli2*. This is a random pattern with respect to *Gli2*, indicating that the 129B7RE marker is not derived from the *Dh* region of chromosome 1 due to the YAC being chimeric. This is in agreement with the inconsistent restriction pattern demonstrated for the right hand end of 129B7.

A *Sty*I polymorphism was detected in 67G3LE; C57BL/6 produced fragments of 230bp, 50bp and 20bp whereas *DBA* yielded fragments of 230bp and 70bp, shown in figure 3.12. The SDP of 67G3LE for the BXD panel is shown in figure 3.2. It is fully concordant with that for *Gli2* demonstrating that the left end of 67G3 is non-chimeric.

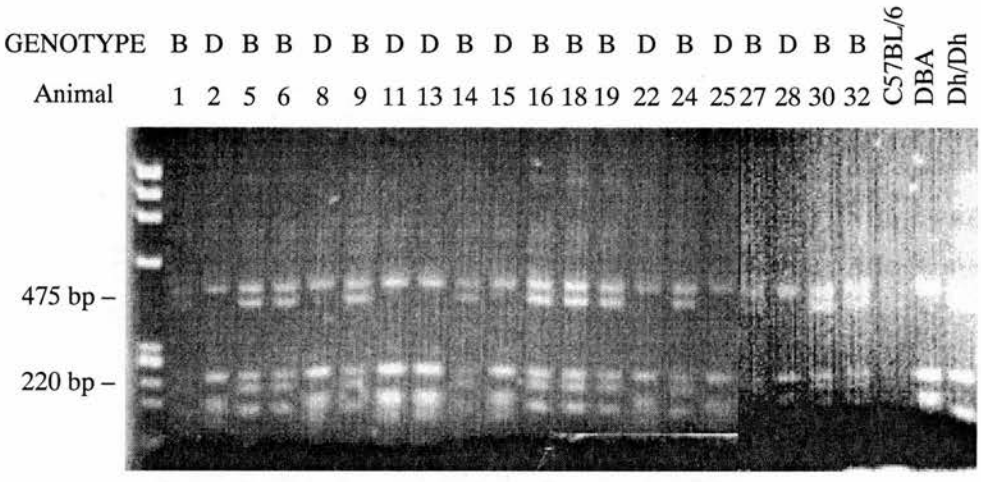



**Figure 3.11 Strain distribution pattern of 129B7RE in the BXD RI strains**

PCR amplification of BXD RI strain DNA using 129B3RE specific primers, H58/H59, followed by *Alu*I digestion. The inheritance of the C57BL/6 allele is determined by the presence of bands of 475bp and 220bp which are absent from the DBA allele. The genotype of each animal is shown.

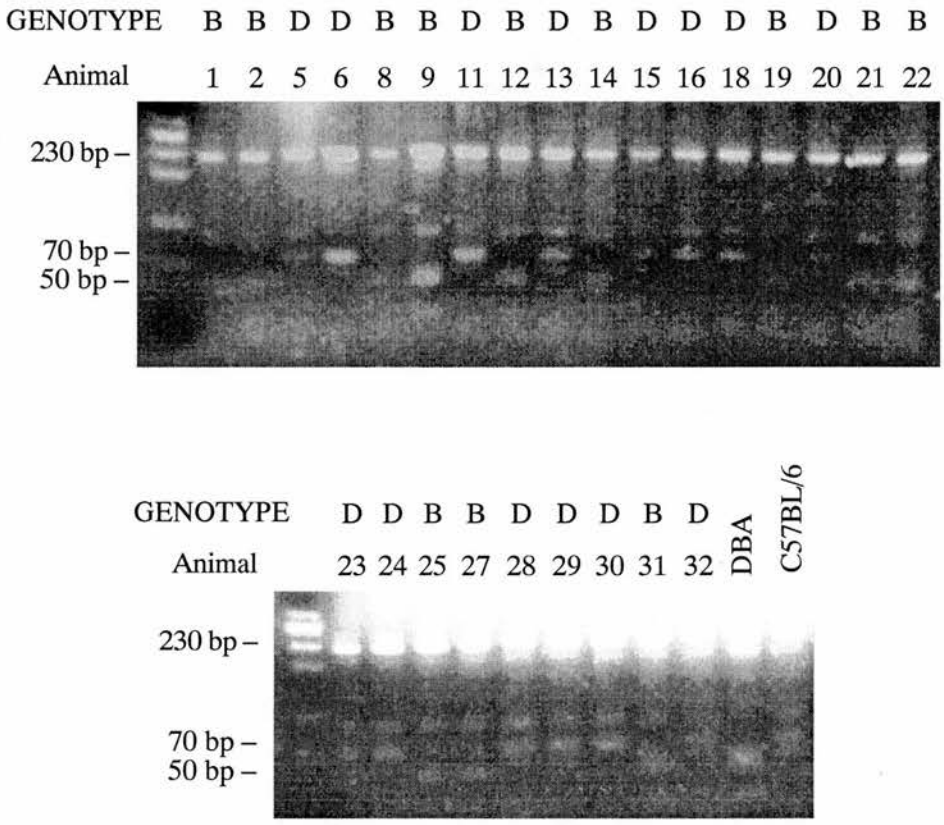



**Figure 3.12 Strain distribution pattern of 67G3LE in the BXD RI strains**

PCR amplification of BXD RI strain DNA using 67G3LE specific primers, I580/I581, followed by *Sty*I digestion. The inheritance of the C57BL/6 allele is determined by the presence of a 70bp compared to a 50bp band in the DBA allele. The genotype of each animal is shown.

### 3.2.6 Placement of YAC derived markers relative to *Dh*

The isolation of YAC clones from the *Dh* region provides a resource for generating markers which can be genetically mapped relative to *Dh*. Only one marker, 67G3LE, yielded a polymorphism which could be mapped on the *Dh* backcross, shown in figure 3.13. The strain distribution pattern produced shows that 67G3LE co-segregates with *Dh*, shown in figure 3.3. This is in agreement with the physical location of 67G3LE, being between the 5′ end of *Gli2* and *Inhβb* (figure 3.17).

### 3.2.7 Physical mapping of YACs

Construction of a YAC contig requires the physical overlap between individual YACs to be established. This was performed by determining the physical restriction map and marker content of each YAC. Restriction analysis depends on the use of rare cutting enzymes to generate a 'ladder' of restriction fragments via partial digestion of YAC DNA which is then resolved by pulsed field gel electrophoresis (PFGE). This allows the ordering of restriction sites with respect to the ends of a YAC by the use of probes specific to each of the YAC vector arms. The order and relative position of restriction sites determined from one end of the YAC can be confirmed by the map generated from the other (Hamvas *et al.*, 1994, Ragoussis, 1996). Within this framework the physical location of specific loci internal to a YAC can be determined.

Such restriction maps allow a detailed comparison of overlapping YACs and enables the detection of deletions or rearrangements and possible chimerism. Further, when available, a comparison can be made to pre-existing maps, either generated from PFGE of genomic DNA or from DNA cloned into other vectors. Additional information can be obtained by the use of rare cutting enzymes which recognise sites located primarily in CpG islands, such as *Not*I, *BssH*II, *Eag*I, (Lindsay and Bird, 1987, Bird, 1989). The location of these sites may represent the position of true CpG islands and thus associated genes (Bird, 1987).

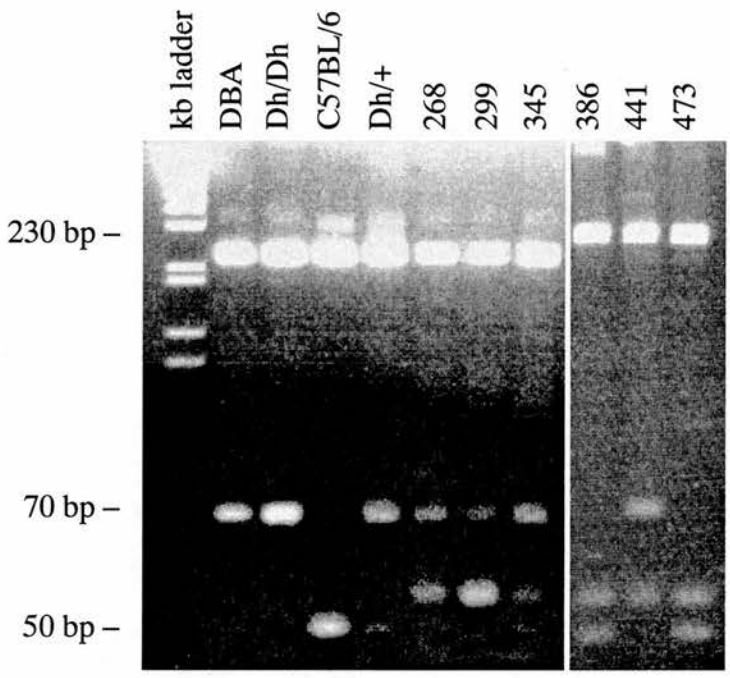


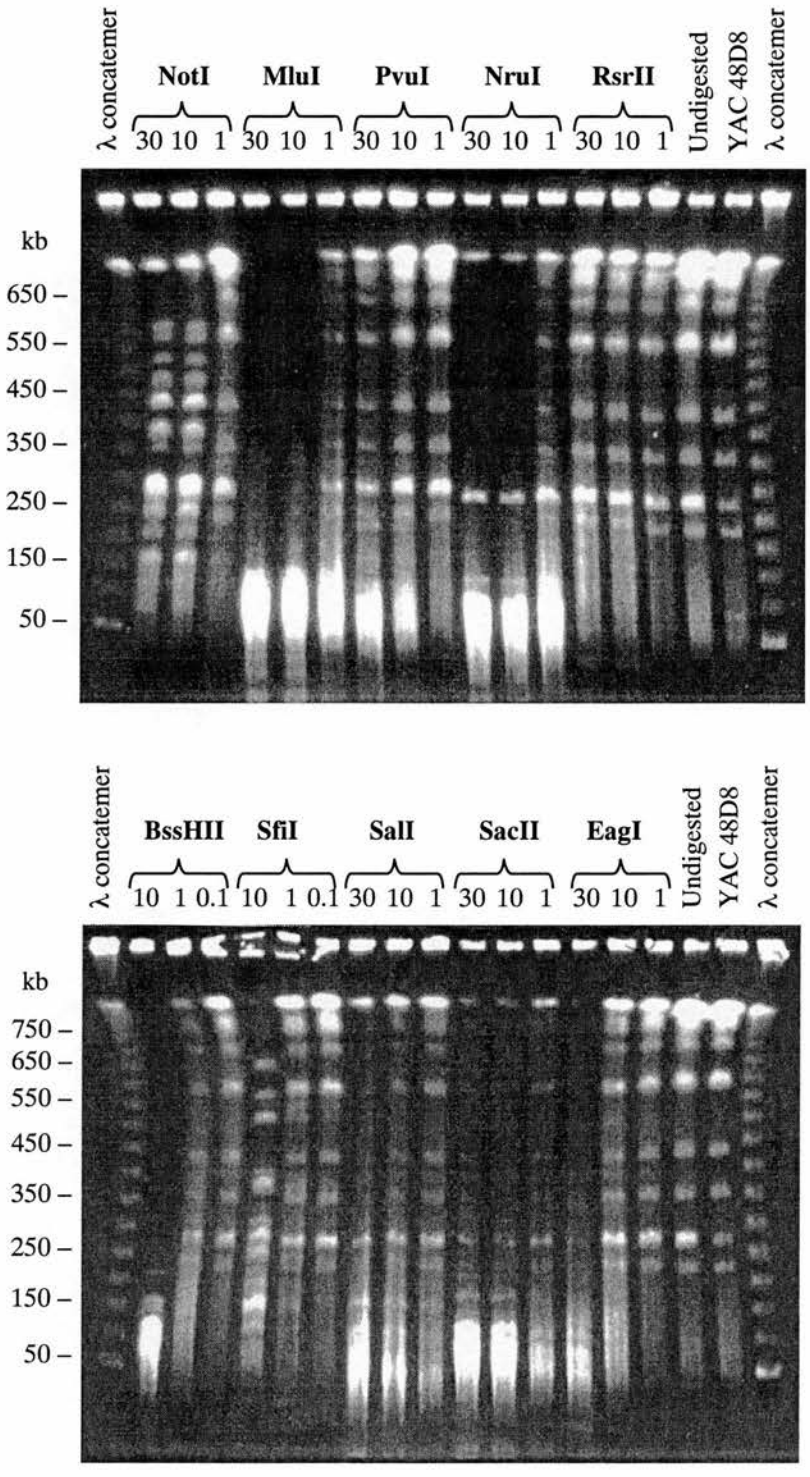

**Figure 3.13 Genetic localisation of 67G3LE relative to *Dh***

Strain distribution pattern of 67G3LE for informative *Dh* backcross progeny. Digestion of the 300bp 580/581 PCR product with *Sty*I produces bands of 230bp, 50bp and 20bp from C57BL/6 and 230bp and 70bp from *Dh* which are separated on a 5% Nusieve agarose gel. (The backcross progeny produce a background band of 55bp).

For each YAC clone, total yeast DNA embedded in agarose plugs was prepared. Initially the size of each YAC was determined. Undigested yeast DNA and lambda 48kb concatamer size markers were separated by PFGE under conditions which resolve between 50Kb and 1Mb. A Southern blot of the gel was then hybridised to lambda marker and YAC vector arm probes, enabling the size of the YAC to be determined relative to the size marker. Additionally this procedure will identify clones containing more than one YAC due to cotransformation and clones containing unstable inserts giving rise to YACs carrying deletions.

Each YAC was then digested with a series of rare cutting enzymes: *Not*I, *Mlu*I, *Pvu*I, *Nru*I, *Rsr*II, *BssH*II, *Sfi*I, *Sal*I and *Sac*II. DNA was digested using three concentrations of enzyme in order to generate partially digested fragments. The DNA was then separated by PFGE using appropriate conditions depending on the size of each YAC (see chapter 2, 2.6.6 and 2.5.4). Southern blots of such gels were then hybridised to a probe specific to one of the YAC vector arms. The blots were then stripped of probe and rehybridised to a probe specific to the other vector arm. The blots were then restripped and hybridised to an appropriate internal probe. This process could be repeated for any additional probes which might be informative.

Examples of this technique are illustrated in figures 3.14-3.16. Figure 3.14 shows ethidium bromide stained gels of YAC 119A4 separated by PFGE following digestion with the panel of rare cutting enzymes at varying concentrations. 119A4 had previously been determined to be 585kb in size enabling these gels to be run under conditions which would separate fragments up to 600kb. The undigested YAC cannot be distinguished as it co-migrates with yeast chromosomes V and VIII, both 585kb. An increase in higher molecular weight DNA can be seen at lower enzyme concentrations indicating successful partial digestion. Figure 3.15 shows blots generated from YAC 20F7 hybridised to the left arm YAC vector probe, LA2.6. Each enzyme can be seen to produce a ladder of bands due to partial digestion of the DNA. The size of the smallest band for each enzyme corresponds to the distance of the nearest restriction site from the right vector arm. The size of the next band


λ concatemer
NotI
MluI
PvuI
NruI
RsrII
30 10 1 30 10 1 30 10 1 30 10 1 30 10 1
Undigested
YAC 48D8
λ concatemer
kb
650 –
550 –
450 –
350 –
250 –
150 –
50 –
λ concatemer
BssHII
SfiI
SalI
SacII
EagI
10 1 0.1 10 1 0.1 30 10 1 30 10 1 30 10 1
Undigested
YAC 48D8
λ concatemer
kb
750 –
650 –
550 –
450 –
350 –
250 –
150 –
50 –

corresponds to the distance of the second restriction site from the first and so on for each band. In this way restriction sites for each enzyme can be sequentially plotted along the YAC. For example *Mlu*I produces five bands at 20kb, 40kb, 120kb, 150kb and 170kb corresponding to *Mlu*I sites at these distances from the left hand vector arm. Similarly for *Sac*II three bands are produced indicating *Sac*II sites 20kb, 200kb and 225kb from the left hand vector arm. Such restriction sites indicated by one vector arm probe are then corroborated by the use of the other arm probe which should identify the same sites.

Figure 3.16 shows blots of the gels shown in figure 3.14 hybridised to the right arm YAC vector probe, RA1.4. A typical pattern of partial restriction fragments can be seen for each enzyme however these blots also illustrate an important characteristic of some YACs. In the lane containing undigested DNA two bands can be seen at 585kb and 400kb. The lower band represents the 119A4 YAC from which 185kb of DNA have been deleted. This is known to be the case as the DNA was prepared from a yeast clone which had previously been shown to contain a single YAC of 585kb. It is therefore most likely that deletion of a portion of the YAC has occurred during propagation of the clone. This phenomenon greatly complicates the interpretation of the data as it can be difficult to determine which bands in the ladder are derived from the intact YAC and which originate from the truncated YAC. Of the 11 YACs analysed six 129B7, 89H7, 119A4, 114C12, 175D3 and 130G11 could not be propagated without the production of truncated YACs. This difficulty could however generally be overcome by mapping from both ends of each YAC and by the restriction pattern produced by internal probes, especially if such probes were deleted in the truncated YAC. Analysis was further aided in regions of overlap between two or more YACs as the restriction map in different YACs should be identical for the region of overlap.

The presence of each end clone marker in all YACs was assessed by PCR analysis. The marker content and restriction patterns generated for each YAC thus

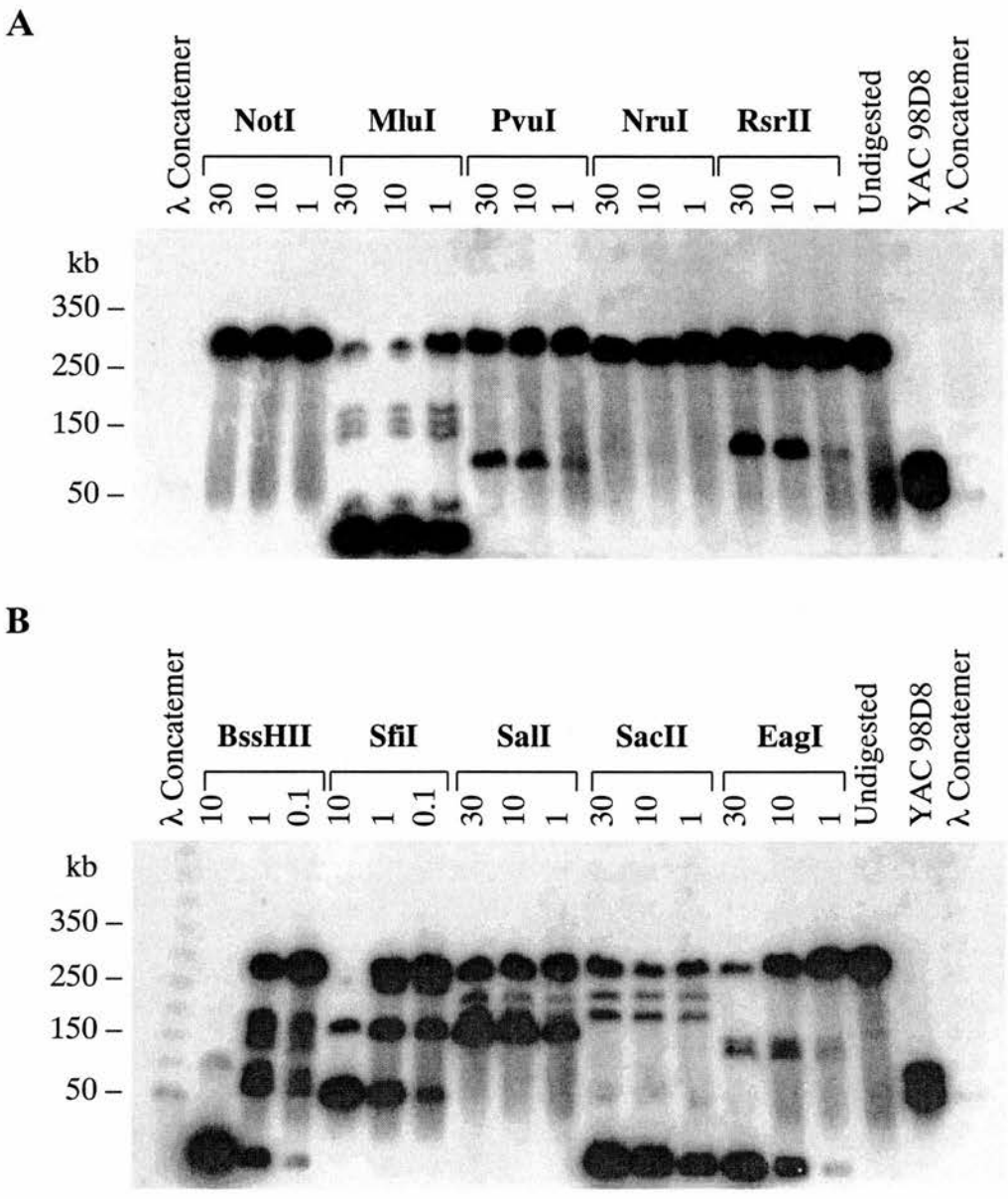



**Figure 3.15 Restriction mapping of YAC 20F7**

YAC 20F7 DNA was restricted using rare cutting restriction enzymes at varying enzyme concentration, 30U/400µl - 1U/400µl, to produce partially digested DNA. DNA was then separated by PFGE under appropriate conditions to separate in the range 0kb - 400kb. Chromosome 5 YAC 98D8 (in the same host yeast strain, AB1380, as all chromosome 1 YACs) was run to act as a control. DNA was then transferred to a nylon membrane and hybridised with the left arm vector probe, LA2.6.

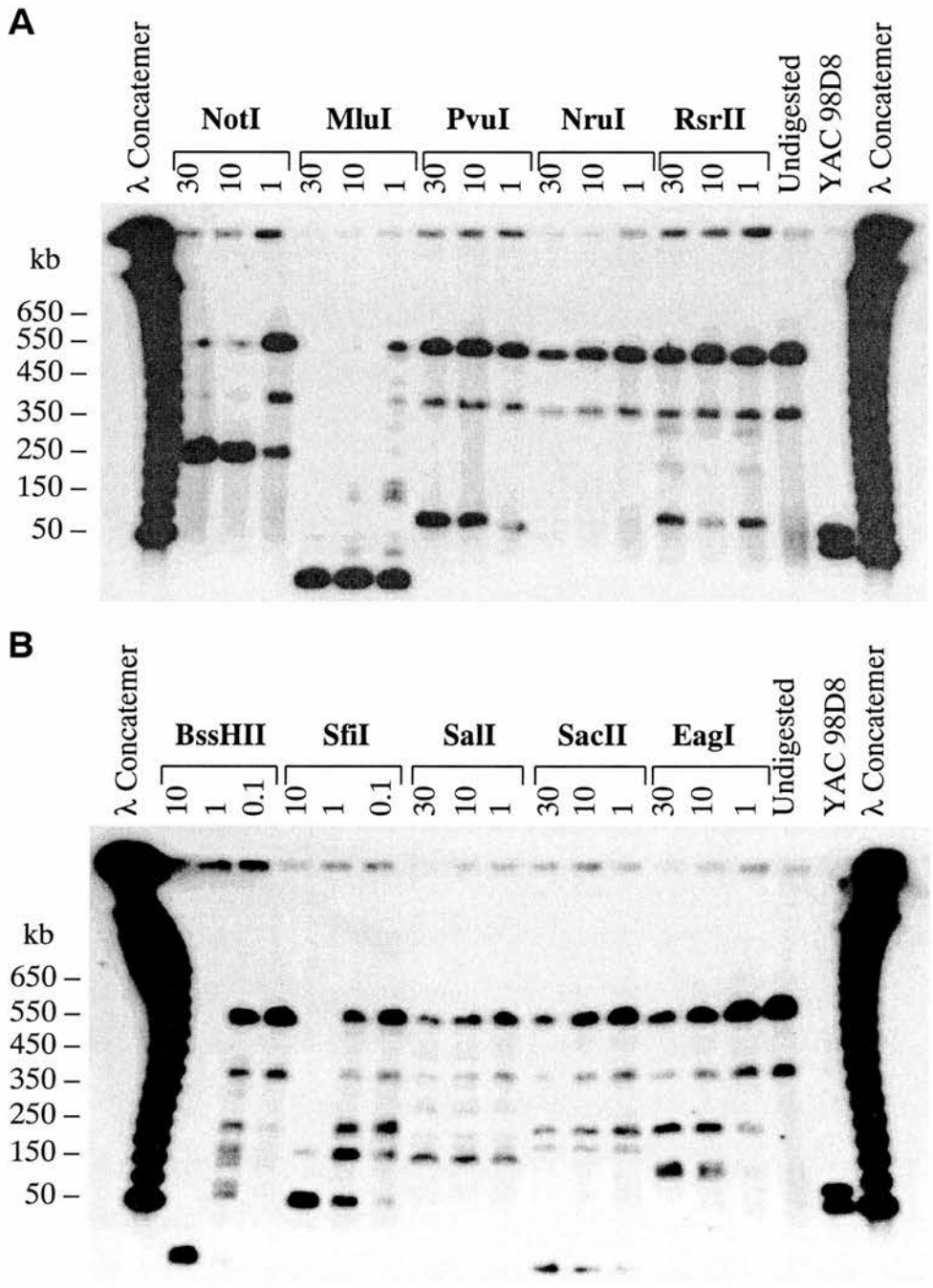



**Figure 3.16 Restriction mapping of YAC 119A4**

YAC 119A4 DNA was restricted using rare cutting restriction enzymes at varying enzyme concentration, 30U/400µl - 1U/400µl, to produce partially digested DNA. DNA was then separated by PFGE under appropriate conditions to separate in the range 0kb - 600kb. Chromosome 5 YAC 98D8 (in the same host yeast strain, AB1380, as all chromosome 1 YACs) was run to act as a control. DNA was then transferred to a nylon membrane and hybridised with the right arm vector probe, RA1.4.

enabled the set of YACs to be aligned relative to eachother. The deduced contig is depicted in figure 3.18

Two YACs, 67G3 and 129B7, were isolated with *Gli2* markers. Since the transcriptional orientation of *Gli2* relative to *Dh* has been shown genetically, restriction mapping of these YACs using 3′ and 5′ *Gli2* probes was performed to determine the orientation of the YACs relative to the chromosome. 5′ and 3′ probes were generated by PCR from *Gli2* cDNA using primers 1a/1b and 5a/5b which amplify the first and last kilobase of coding sequence respectively. The entire coding sequence is demonstrated to lie in a region of 80kb, flanked by *Eag*I and *Sal*I sites 5′ and a SacII site 3′. This identified the left end of 67G3 and the right end of 129B7 as extending distally from *Gli2*, into the *Dh* region.

The restriction sites proximal to *Gli2* in 67G3 are uncorroborated and may not represent the endogenous genomic region. This is because the right arm probe hybridises to a 67G3 YAC of 475kb, while the left hand probe and the *Gli2* probes recognise a YAC of 1Mb. This indicates a rearrangement resulting in the YAC vector arms being located on different DNA fragments. The region distal to *Gli2* represents the *Dh* region because it is corroborated by YACs 129B7, 119A4, 76B3 89H7 and 20F7 (although there is an apparent deletion in 67G3), see below. Also the left end clone of 67G3 is demonstrated genetically to reside in the *Dh* region. Since the YAC has undergone rearrangement, the map proximal to *Gli2* may therefore be non-contiguous as it is largely uncorroborated. There is however a single corresponding *Eag*I site between 67G3 and the genomic map in this region. *BssH*II and *Mlu*I sites are also placed on the genomic map in this region but because it was not possible to map using the right vector arm probe it could not be determined if these sites were present in the YAC.

The restriction sites 100kb 5′ of *Gli2* concur between 67G3 and 129B7 and thereafter diverge. This is due to the distal region of 129B7 being chimeric, as shown by the genetic localisation of the right end clone marker on the BXD RI strains and a

discordant restriction pattern compared with overlapping YACs 119A4, 76B3 and 89H7. 67G3LE however is shown to be contiguous due to cosegregation with *Gli2* on the BXD RI strains and with *Dh* on the *Dh* backcross. Three of the four YACs, 119A4, 76B3, 89H7 and 20F7, subsequently isolated with this marker, display identical restriction maps and end clone marker content in the regions of overlap. They also contain *Inhβb* which, by analysis of the BXD RI strains and the *Dh* backcross, was also shown to reside within the *Dh* region. In addition those restriction sites determined from the genomic map are all present in these YACs. Together this demonstrates no sign of chimerism or rearrangement.

The fourth YAC, 20F7, does however exhibit either chimerism or a large deletion at the proximal end. Alignment of the map for 20F7 shows the restriction pattern of the proximal most 75kb does not concur with the consensus. In addition hybridisation and PCR analysis fail to detect the presence of *Inhβb*. For 67G3LE to be present in 20F7 it must lie distal to *Inhβb*, but by no more than 60kb because it is also present in 76B3. Two sites, *Sal*I and *Sac*II, adjacent to 67G3LE are seen in the four overlapping YACs, but the *Mlu*I and *Sal*I sites either side of *SalI*/*Sac*II are not present in 67G3. This indicates 67G3LE to lie distal to the *Sac*I/*Sal*I sites but proximal to the *Mlu*I site in 119A4, 76B3, 89H7and 20F7. Clones 119A4, 76B3 and 89H7 extend to a maximum of 375kb proximal to *Inhβb* and the right end clone of 89H7 is present in 67G3 while the left end clone of 119A4, which is 75kb distal to that of 89H7 is not (presence of the left end of 76B3 in 67G3 has not been determined). The absence of *Inhβb* and 119A4LE from 67G3 and the incompatible restriction maps of 67G3 compared to 119A4, 76B3 and 89H7 therefore indicates a deletion in 67G3. This deletion is estimated to extend from a point 100kb distal to *Gli2*, distal to the *Rsr*II site because proximal to this the maps of 67G3 and 129B7 concur (except for an *Eag*I site seen in 67G3 where a *Not*I site is seen in 129B7. As the recognition site of *Eag*I is internal to that of *Not*I this could be due to polymorphism as 67G3 is of C57BL/6 origin while 129B7 is from the C3H strain).

In 67G3, 89H7RE must be located in a region of 50kb between this *Rsr*II site and the deletion breakpoint because the *Rsr*II site is absent from 89H7 and the most proximal site in 89H7, *BssH*II, is 50 kb from 89H7LE and is absent from 67G3. The deletion extends up to a point between the *Sal*I site 25kb distal to *Inhβb* and the *Sal*I/*Sac*II sites 50kb distal to *Inhβb* and is therefore estimated to be between 350kb-400kb. 76B3 and 89H7 extend proximally 375kb but neither restriction map overlaps with that of 67G3. The map distal to *Gli2* extends 100kb before the deletion and the window of overlap between 89H7 and 67G3 is 50kb. The distance between *Gli2* and *Inhβb* is therefore 475kb - 525kb. The YAC contig extends distally from *Inhβb* to a maximum of 275kb therefore the total extent of the contig from the 3′ of Gli-2 is 830kb - 880kb.

Three YACs containing *En1* have been isolated which extend a maximum of 650kb to the 5′ of *En1* and 325kb 3′, however no overlap with the *Inhβb*/*Gli2* YACs can be detected. There is no overlap between the restriction maps of the two sets of YACs and end clone markers were found not to overlap by hybridisation and PCR analysis. The transcriptional orientation of *En1* relative to the chromosome is not known, it is therefore not known to which side of the *En1* contig the *Dh* region extends.

The map of 130G11, which extends 650kb to the 5′ of *En1*, corroborates with that of 114C12, which is judged to be non-chimeric by FISH analysis (see section 3.2.5), while both contain the sites present on the genomic map (although *Eag*I, placed on the genomic map, was not mapped in these YACs). The microsatelite marker, *D1Mit*390, shown by PCR analysis to be uniquely present in 130G11 was determined to be located to a 25kb region 550kb upstream of *En1*. Generation of a map 5′ of *En1* in 175D3 proved difficult due to the complication of YAC truncation during propagation, allowing only four sites to be mapped. However absence of *D1Mit*390 suggests the clone is either chimeric or rearranged. Again due to YAC truncation it was not possible to accurately map all sites in the 325kb region 3′ of

*En1* in 130G11. There are therefore some sites which could not be positively placed which are present in the overlapping region of 175D3. The left end clone of 175D3 is present in 130G11 and common *Sfi*I and *Sal*I sites are seen which are also present in the genomic map, suggesting these regions to be contiguous.

Two further YACs, 97A1 (800kb) and 72C4 (300kb), have been isolated using the left end clone of 175D3. Both contain the right end clone of 130G11 but neither contain *En1* or other end clone markers. They have not been analysed further.

### 3.2.8 Generation of a genomic map of the *Dh* region

The construction of a long range physical map of the *Dh* region was attempted by the analysis of genomic mouse DNA. High molecular weight DNA was prepared in agarose plugs from C57BL/6 liver tissue (a kind gift from Andreas Schedel), digested with a series of rare cutting enzymes and separated by PFGE to resolve fragments between 50kb to 1Mb. Gels were then transferred to nylon membranes and sequentially hybridised to the available probes from the region.

Probes were prepared from the 5′ of *Gli2* and from the 3′ and 5′ regions of the *En1* and *Inhβb* genes. Both 3′ and 5′ probes were prepared because of the identification of CpG islands associated with *En1* and *Inhβb*. If 5′ and 3′ probes hybridise to different sized fragments for a particular enzyme then the restriction site must lie between the probes. This enables the identification of fragments extending both 5′ and 3′ from the gene and provides an anchor for the positioning of fragments within the map.

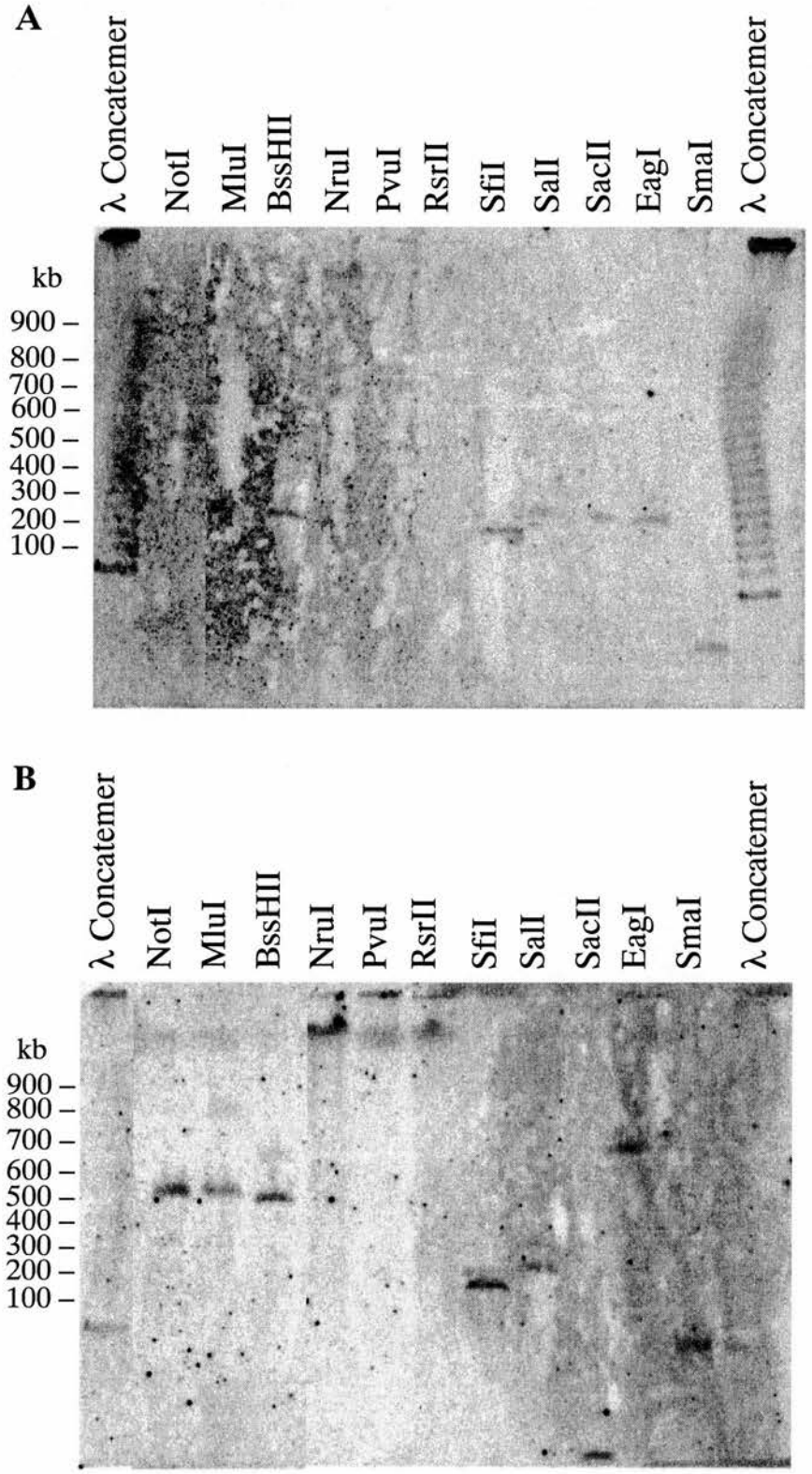
The 5′ *En1* probe was a 1.2kb *Sal*I/*Mlu*I fragment situated 3kb 5′ to *En1*, (Higgins, 1991) while the 3′ probe was a 500bp cDNA fragment, pλ4L7, located 3′ of the homeobox (Davidson *et al.*, 1988). The 5′ *Gli2* probe was the most 5′ 1.3kb of coding sequence generated by PCR amplification of *Gli2* cDNA using primers 1a/1b. The 3′ *Inhβb* probe was also generated by PCR amplification of a 671bp fragment from the last exon of the gene using primers I181/I182.

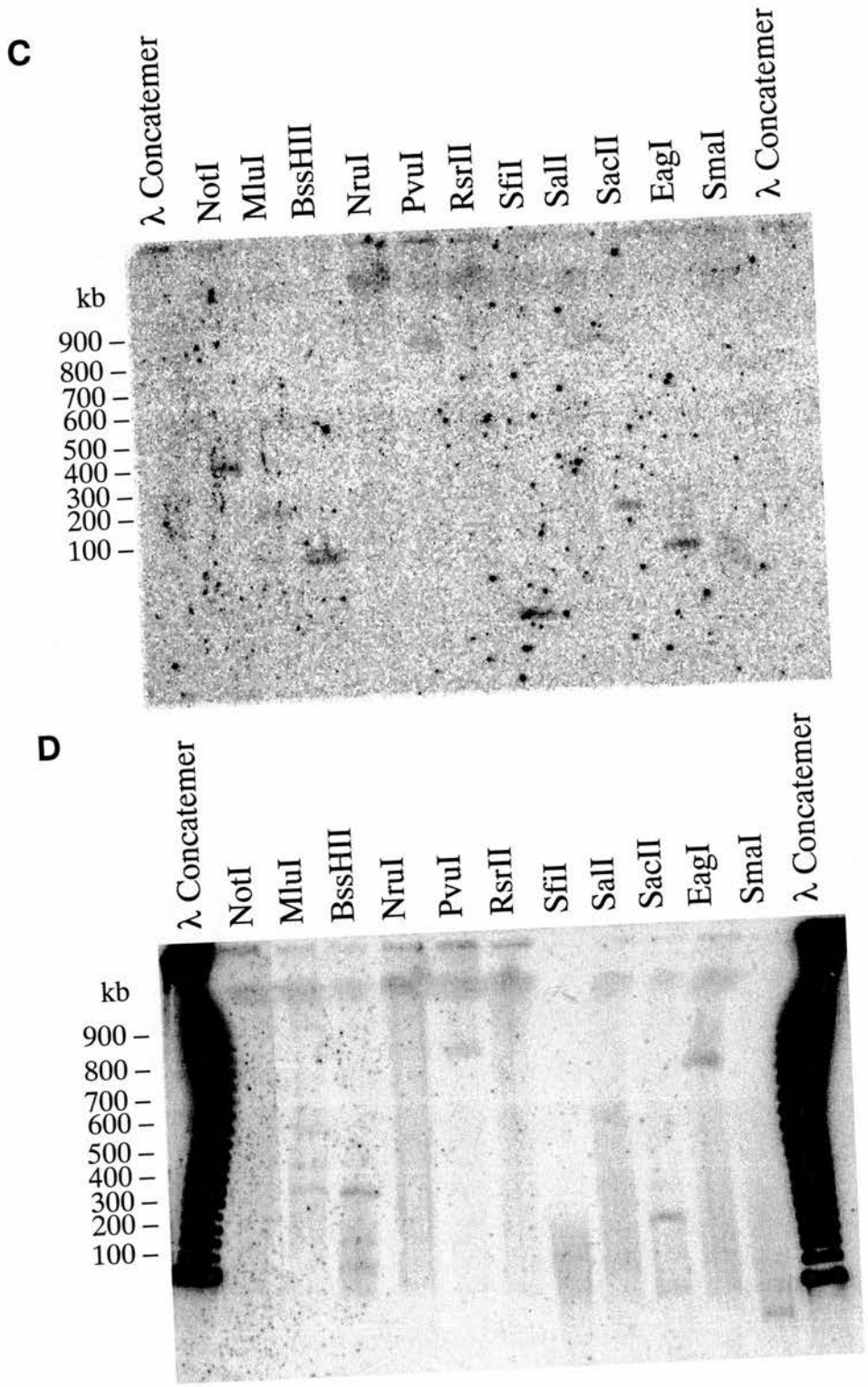
The 5′ *Inhβb* probe was isolated from the *Inhβb* lambda clone described previously, (3.2.1). Restriction mapping of the lambda clone identified a single *Not*I site flanked by 4.0kb and 9.0kb of DNA. The 3′ *Inhβb* probe hybridised to the 9.0kb fragment only. Assuming the *Not*I site to be located within a CpG island at the 5′ end of *Inhβb*, then the 4.0kb fragment would reside 5′ to the gene. To identify a single copy fragment, for use as a probe, the 4.0kb fragment was subcloned into pBluescript and digested with a series of enzymes to identify fragments of approximately 1kb. *Sac*II produced fragments of 1.8kb, 1.2kb, 700bp and 400bp. These fragments were assayed by hybridisation to Southern blots of mouse genomic DNA. The 1.2kb fragment was judged to be single copy and thus suitable as an *Inhβb* 5′ probe.

Figure 3.17 shows these five probes hybridised sequentially to the same nylon filter. Between each hybridisation the filter was striped of probe and checked for absence of signal by phosphorimage analysis. The map derived from this data is illustrated in figure 3.18. For all digests, except *Pvu*I the *Inhβb* 3′ and 5′ probes hybridise to different fragments demonstrating the presence of a CpG island situated between the probes. The 3′ *Inhβb* probe hybridises to three *BssH*II fragments of 75kb, 175kb and 350kb and to two *Mlu*I fragments of 175kb and 350kb. This is most likely due to partial digestion caused by partial methylation at these sites. This probe also hybridises to an 825kb *Eag*I fragment and an 875kb *Pvu*I fragment. The physical linkage of 475kb - 525kb demonstrated on the YAC contig between *Gli2* and *Inhβb* suggests that the *Gli2* 5′ probe should also hybridise to these fragments. Due to high background hybridisation the 825kb *Eag*I fragment cannot be discerned with the *Gli2* 5′ probe but the same 875kb *Pvu*I fragment is detected. However the *Inhβb* 5′ probe also hybridises to this *Pvu*I fragment preventing the relative placement of the *Pvu*I sites on the map. The *Gli2* 5′ probe recognises different sized *Mlu*I and *BssH*II fragments to *Inhβb* 3′. The location of the proximal sites for these fragments are

placed on the map assuming that the distal sites are those detected by the *Inhβb* 3′ probe.

The *En1* 3′ and 5′ probes also detect different sized fragments demonstrating the CpG island known to reside at the 5′ of *En1* (Higgins, 1991, Logan *et al.*, 1989). No common fragments are detected by the *Inhβb* 5′ probe and either *En1* probe; therefore no linkage between the two is achieved. The minimum physical linkage, if the 5′ of *En1* lies proximally (as depicted in figure 3.17), is 1.05Mb as deduced from the YACs 130G11 and 119A4, 89H7 and 20F7. If *En1* is in the opposite orientation the minimum physical linkage is 1.15Mb deduced from the 400kb *Inhβb* 5′ *Not*I fragment and the 750kb *Eag*I *En1* 3′ fragment.


A
λ Concatemer
NotI
MluI
BssHII
NruI
PvuI
RsrII
SfiI
SalI
SacII
EagI
SmaI
λ Concatemer
kb
900 –
800 –
700 –
600 –
500 –
400 –
300 –
200 –
100 –
B
λ Concatemer
NotI
MluI
BssHII
NruI
PvuI
RsrII
SfiI
SalI
SacII
EagI
SmaI
λ Concatemer
kb
900 –
800 –
700 –
600 –
500 –
400 –
300 –
200 –
100 –


C
λ Concatemer
NotI
MluI
BssHII
NruI
PvuI
RsrII
SfiI
SalI
SacII
EagI
SmaI
λ Concatemer
kb
900 –
800 –
700 –
600 –
500 –
400 –
300 –
200 –
100 –
D
λ Concatemer
NotI
MluI
BssHII
NruI
PvuI
RsrII
SfiI
SalI
SacII
EagI
SmaI
λ Concatemer
kb
900 –
800 –
700 –
600 –
500 –
400 –
300 –
200 –
100 –

**E**

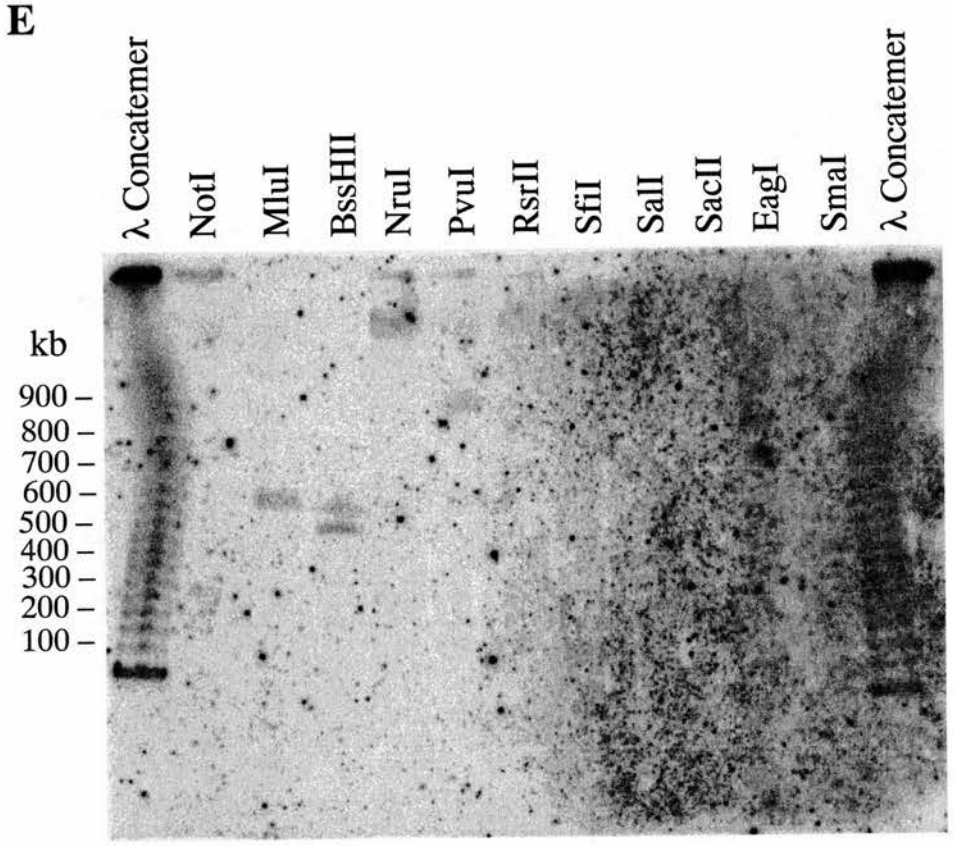
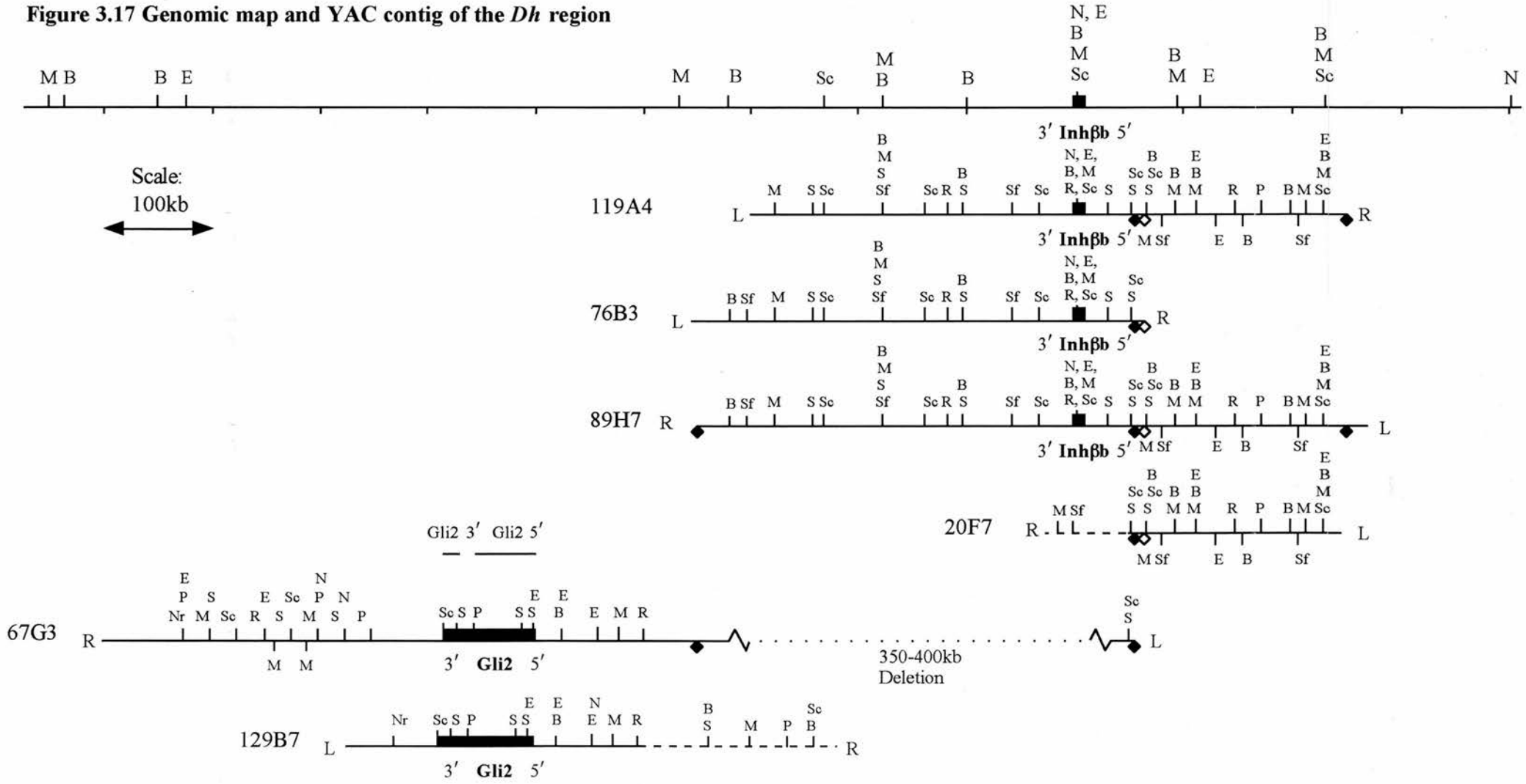
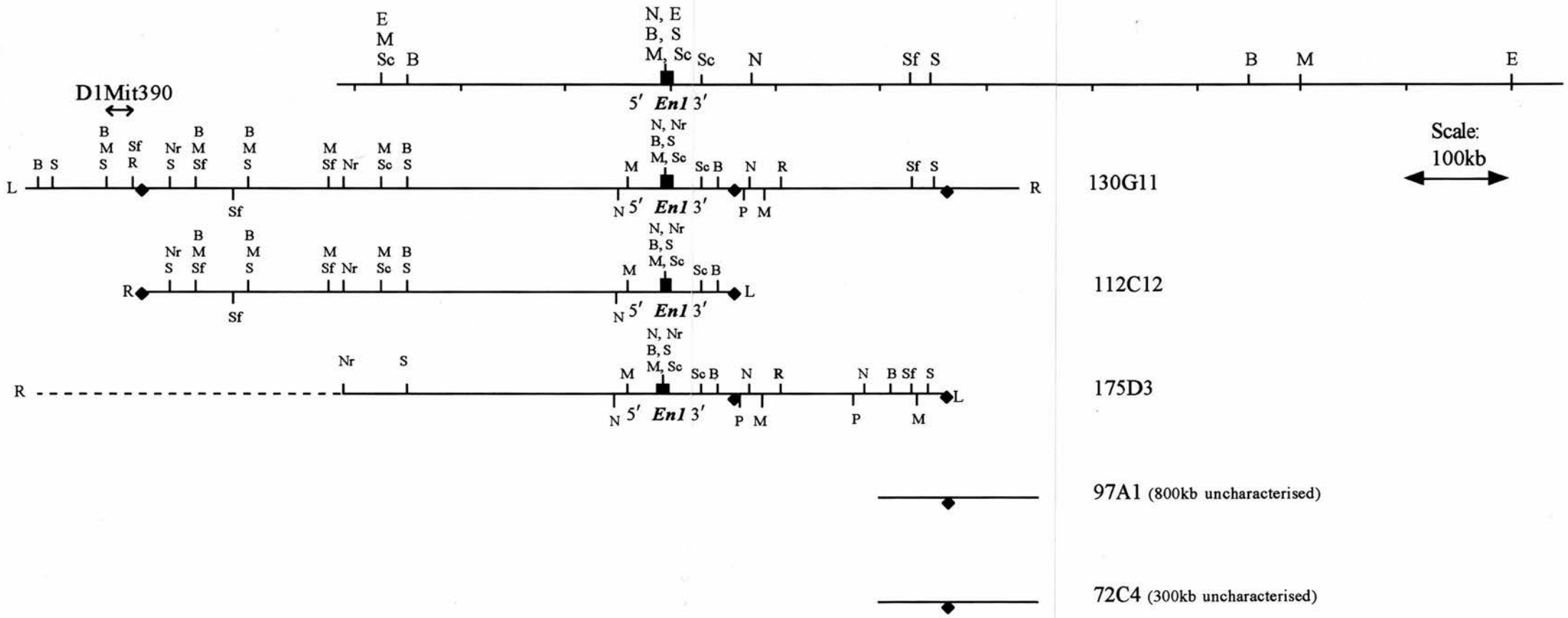

λ Concatemer
NotI
MluI
BssHII
NruI
PvuI
RsrII
SfiI
SalI
SacII
EagI
SmaI
λ Concatemer
kb
900 –
800 –
700 –
600 –
500 –
400 –
300 –
200 –
100 –

**Figure 3.17 Genomic map and YAC contig of the *Dh* region**




D1Mit390
5′ En1 3′
Scale:
100kb
130G11
112C12
175D3
97A1 (800kb uncharacterised)
72C4 (300kb uncharacterised)

### 3.3 Discussion

A fundamental prerequisite to understanding the developmental basis of the *Dh* phenotype and the role the *Dh* gene plays during normal development is the identification of the *Dh* gene and the molecular nature of the mutation. The genetic localisation of *Dh* to a defined interval of 1.2cM provides the starting point for the molecular cloning and characterisation of this genomic region. This in turn provides the resource to identify candidate genes for *Dh*. Since a genetic region of 1.2 cM is likely to be in excess of 2Mb, cloning of the *Dh* region was approached by the construction of a YAC contig.

Two YACs, 67G3 and 129B7, were isolated with *Gli2* markers, the 3′ untranslated sequence of which marks the proximal extent of the *Dh* region. Orientation of *Gli2* within these YACs then enabled the left end clone of 67G3 to be used to isolate four overlapping YACs 119A4, 76B3, 89H7 and 20F7, which extend further into the *Dh* region. The PFGE restriction map and marker content of each clone was determined enabling YACs to be aligned relative one another. The contig is presumed to present a faithful representation of the genomic region because of the overlap detailed between the YACs and because of the exact corroboration observed between the genomic map and the contig. However because of the deletion present in 67G3 the exact degree of overlap between YACs 67G3 and 89H7 cannot be established but is judged to be within a maximum region of 50kb. The contig extends 830kb - 880kb from the 3′ of *Gli2* into the *Dh* region.

Within this region two genes, *Gli2* and *Inhβb*, have been localised. The *Gli2* gene produces a message of 8.0kb (Hui *et al.*, 1994), 4.6kb of which is coding. Using the first and the last kilobase of coding sequence as 5′ and 3′ probes respectively the coding region of the *Gli2* gene is located to an 80kb region. This was somewhat surprising because the human *GLI3* locus has been shown to span 280kb (Vortkamp *et al.*, 1994; Vortkamp *et al.*, 1995). *GLI3* is the most homologous gene to *Gli2*, 66% homology is seen at the nucleic acid level (Hughes *et al.*, 1997). This very high degree

of homology suggests the two genes have arisen by duplication from an ancestral *Gli* gene. One would therefore expect the genomic organisation of the two genes also to be similar. Comparison of the genomic organisation of *GLI3* and the YAC contig however suggests that the organisation of the two genes are indeed similar. Vortkamp *et al*, (1995) have generated a cosmid contig of the *GLI3* genomic region. They demonstrate the distal 3.3kb of *GLI3* cDNA to reside in only 30kb while the proximal 1.7kb of cDNA is spread over a region of 250kb. The 3′ *Gli2* probe locates to a region of approximately 10kb while the 5′ *Gli2* probe hybridises to exons in the adjacent 70kb region. If the 5′ exons of *Gli2* are distributed over a long range, as in *GLI3*, then they will not be present in either 67G3 or 129B7 because these YACs are not contiguous for more than 100kb beyond the detected 5′ of *Gli2*. If this is the case then failure to detect *Gli2* coding sequence in this 100kb indicates an intron of at least this size. This compares to the 5′ of *GLI3* in which there are two large introns of 60kb. The overlap between 67G3 and 89H7, 76B3 and 119A4 would predict that the *Gli2* 5′ probe should detect the presence of 5′ exons in these YACs, however this has not been assessed. Recently however a genomic lambda clone has been isolated by Carlo De Angelis which contains a region of the *Gli2* 5′ untranslated sequence. A probe derived from this lambda clone does not hybridise to 67G3 or 129B7 but does hybridise to 89H7, 76B3 and 119A4, although the specific location has not been determined. This confirms the *Gli2* gene to cover at least 280kb, the distance from the 3′ of *Gli2* to YAC 119A4. The genomic organisation of *Gli2* is therefore likely to be very similar to *GLI3*. Vortkamp *et al* (1995) report a CpG island for *GLI3*; however, there is not obviously an island associated with *Gli2*.

*Inhβb* was demonstrated to cosegregate with *Dh* in the *Dh* backcross. Concomitantly *Inhβb* was determined to reside within YACs 119A4, 76B3 and 89H7. The restriction analysis of these YACs combined with the genomic map places *Inhβb* 550kb - 600kb distal to the 5′ of *Gli2*. With the genetic resolution provided through the *Dh* backcross this represents a minimum distance in which *Dh*

can lie because *Gli2* and *Inhβb* are non-recombinant with *Dh*. *Inhβb* is a subunit of both activin and inhibin, dimeric growth factors of the *TGFβ* superfamily, which are a class of peptides known to regulate growth and differentiation in a variety of cell types. The chromosomal location and biological nature therefore identify *Inhβb* as a candidate gene for *Dh*, investigated further in chapter four.

It is possible to speculate on the location of *Dh* relative to *Inhβb* because of paralogous linkage groups present on chromosomes 1, 5 and 13. The *Dh* gene has been predicted to be homologous to the *Hx* gene because the loci are closely linked to the genes *En1-Gbx2-Htr5b* on chromosome 1 and *En2-Gbx1-Htr5a* on chromosome 5 respectively. This paralogous linkage and the phenotypic similarities of the two mutations suggest the *Dh* and *Hx* genes arose from the duplication of a chromosomal segment containing the ancestral *Dh/Hx* gene and ancestral *En*, *Gbx* and *Hrt5* genes (this is fully discussed in chapter 5). However the linkage of *Inhβb* to *Gli2* indicates paralogy with chromosome 13 because *Gli3* is tightly linked to *Inhβa*. This suggests an extension of the original duplication followed by rearrangement resulting in the separation of *Gli3/Inhβa* from *Hx/En2/Gbx-1*. That *Hx* is linked to *En2* and not *Inhβa* suggests that *Hx* must have originally been located on the *En2* side of *Inhβa* and not the *Gli3* side. Thus if *Dh* and *Hx* are truly paralogous genes *Dh* would be expected to lie on the *En1* side of *Inhβb*, not on the *Gli2* side.

Three YACs constitute the contig based around *En1* which extends upstream from *En1* by 625kb and downstream by 325kb. No overlap between the *En1* contig and the *Gli2/Inhβb* contig can be established. Similarly no linkage is observed on the genomic map. Presently therefore the physical extent of the *Dh* region, between *Gli2* 3′ and *En1*, is 1.45 - 1.5Mb if the 5′ of *En1* is proximal or 1.7 - 1.75Mb if the 5′ of *En1* is distal. This compares to the 1.2 cM *Dh* interval which on average represents 2.4Mb.

Construction of the contig can continue on a number of fronts. The most distal end clone markers, 119A4RE, 89H7LE and 20F7LE can be used to rescreen the

YAC libraries in order to identify additional clones which will extend further into the *Dh* region. These YACs can then be assessed for crossover into the *En1* YACs. If no crossover is observed then appropriate end clone markers can be isolated and used to perform additional walking steps. The same strategy should of course be followed on the *En1* side of the contig. However a crucial point to establish is the orientation of the *En1* contig relative to the chromosome to enable the contig walk to proceed in the correct direction. This can be achieved either through the establishment of physical overlap between the *En1* and *Gli2/ Inhβb* contigs or by the physical localisation of a marker on the *En1* contig which is known to lie either proximal or distal to *En1*. Two YACs 97A1 and 72C4 were isolated using 175D3LE which should extend downstream of *En1*. Although these YACs have not been characterised neither contain any markers present in the *Gli2/Inhβb* contig.

The orientation of *En1* can be assessed either by generating YAC derived markers which are then genetically mapped relative to *En1* on appropriate crosses. Alternatively markers which have previously been shown to genetically map in the vicinity of *En1* can be physically located on the YAC contig. Polymorphic markers were sought between C57BL/6 and DBA or the *Dh* chromosome for 130G11LE, 130G11RE and 175D3LE by restriction analysis (see 3.2.5) but none were detected. This would have enabled typing of the BXD RI strains and the three animals which are recombinant between *Dh* and *En1*. Alternative methods could be used to detect polymorphism, for example the end clones could be used to probe Southern blots of C57BL/6 and *Dh/Dh* DNA restricted with a range of enzymes. Also the end clones could be used to isolate genomic lambda clones from which polymorphic microsatellite repeats could be detected, as was performed for *Inhβb* (section 3.2.1). Alternatively variant loci could be sought between C57BL/6 and *Mus spretus* to enable the genetic order of markers to be assessed on the European Interspecific collaborative backcross (EUCIB) (EUCIB group, 1994). This would have two advantages; firstly variant loci between C57BL/6 and *Mus spretus* are much more readily detected than polymorphism between two inbred strains of the same species

(Love et al., 1990; Dietrich et al., 1992). Secondly the average genetic resolution of EUCIB is 0.1 cM which increases the likelihood of being able to order markers.

*D1Mit* markers were assessed for their map position relative to *En1* on the *Dh* backcross. Unfortunately markers which mapped either side of *En1* were not located on the contig. *D1Mit*390 was however localised 525kb upstream of *En1* in 130G11. Although *D1Mit390* cosegregates with *En1* on the *Dh* intraspecific outcross the relative marker order may be discerned on EUCIB. If *D1Mit*390 is shown to be proximal to *En1* then it would define a new minimal region for *Dh*.

New genes and markers which arise in the literature and map in the vicinity of *Dh* or to the syntenic human region, 2qcen-*2*q21, should be tested for their segregation with *Dh*. As with *Inhβb*, the YACs can then be tested for the presence of such a marker. Should the marker not be present in the contig it will be located in a region not represented and can therefore be used to screen for additional YACs. In addition, if a suitable probe can be generated, such markers can be added to the genomic map. If a gene with biological significance to *Dh* is shown to be non-recombinant with *Dh* then it can immediately be assessed for mutation in *Dh* animals.

The genetic interval in which *Dh* resides can be refined because three animals are recombinant between *Dh* and *En1*. The localisation of markers which are proximal to *En1* and which segregate with *Dh* will redefine the distal extent of the *Dh* region (as discussed for *D1Mit*390). It is important to assay markers generated from the *Gli2/Inhβb* contig for segregation with *Dh*. The YAC end clones 119A4RE, 89H7LE and 20F7LE should be analysed to generate such a polymorphic marker. Should one of these lie on the distal or *En1* side of any of the three recombinants then it would delimit *Dh* to the *Gli2/Inhβb* contig. Similarly any other distal markers subsequently isolated, such as end clones from additional YACs, should be assayed in this way.

The *Dh* YAC contig provides the resource to identify candidate genes for *Dh*. Apart from *Gli2* and *Inhβb* the PFGE map indicates the location of several genes by association to CpG islands. CpG islands, usually 1-2kb long, are characterised by a

high density of non-methylated CpG dinucleotide, 10-20 times above that found in bulk DNA, (Bird, 1986; Larsen *et al.*, 1992). This apparent high density of CpG in island sites is due to a depletion of CpG in the rest of the genome where CpG is present at less than a quarter of the expected frequency (Bird *et al.*, 1985; Tykocinski *et al.*, 1984). This lack of CpGs in bulk DNA is thought to be due to the consequence of cytosine methylation. 5-methylcytosine is prone to deamination resulting in conversion to thymine. The CpGs at island sites do not become methylated and are therefore not prone to this conversion.

It is clear that the vast majority of CpG islands are located at the 5′ of genes, particularly in the region of transcription initiation and therefore represent important molecular markers for genes (Bird, 1986). Larson *et al.*, (1992) screened 375 genes in the Genbank database, finding virtually all widely expressed housekeeping genes and 40% of tissue specific genes are associated with CpG islands. Further of the tissue specific genes which lack CpG islands many have shorter stretches (<100bp) of a (G+C) rich sequence. Antequera and Bird, (1993) similarly show 56% of all genes associated with CpG islands. The restriction enzymes used to generate PFGE maps by necessity cleave DNA infrequently. This property is because these 'rare-cutting' enzymes all contain CpG, usually two copies, in their recognition sequence and are sensitive to cytosine methylation. Therefore, in uncloned DNA, rare-cutting enzymes cleave DNA primarily at CpG islands. Lindsay and Bird, (1987) and Bird, (1987) calculate 89% of *Not*I sites and 74% of *Eag*I, *Sac*II and *BssH*II sites to be present in CpG islands. Island sites for other rare cutting enzymes, for example, *Mlu*I and *Pvu*I are predicted to represent a smaller proportion, 27%, of the total. These enzyme sites are therefore seen to cluster at CpG islands and are thus readily detectable on PFGE restriction maps. Because cloned DNA is propagated in bacterial or yeast hosts DNA methylation is lost. Restriction analysis of cloned DNA therefore detects additional sites compared to maps derived from genomic DNA. This can be clearly seen in figure 3.18 where many sites have been mapped in the YACs but which, due to methylation, are not detected in genomic DNA.

The two most predominant clusters of sites seen in the *Dh* contig, (figure 3.18), occur at the *Inhβb* and *En1* loci where sites for *Not*I, *Eag*I, *BssH*II, and *Sac*II are seen. A CpG island has previously been reported for *En1* (Logan *et al.*, 1992; Higgins 1991). The 5′ of *Gli2* is predicted to be 280kb proximal to the 3′ limit of the gene (see above). The genomic map demonstrates a *BssH*II site 250kb and a *Sac*II 350kb from the 3′ of *Gli2*, while no additional sites are detected in the YACs. No obvious CpG island is therefore present in this region indicating *Gli2* does not possess a CpG island. The other clusters of sites which may indicate true CpG islands, indicated on figure 3.18, are 175kb proximal to *Inhβb*, 75kb, 100kb and 225kb distal to *Inhβb* and 275kb, 400kb, 450kb and 550kb upstream of *En1*. Unclustered *Not*I, *Eag*I, *BssH*II, and *Sac*II sites detected in the genomic map may also indicate CpG islands, because of their high percentage association. These plus the clustered sites total 21 possible islands in a region of 2.3Mb.

The marking of genes by CpG islands presents useful information to aid the cloning of genes from large genomic clones. For example those clones, either YACs or YACs subcloned into cosmids or lambda vectors, containing CpG islands can be preferentially analysed for coding sequences. More directly, because the enzymes which cut in CpG islands are 'rare cutters' they can be exploited to subclone island site fragments into plasmid vectors in a conventional manner. Also Valdes *et al.*, (1994) describe a vectorette based PCR methodology to amplify sequences at CpG islands directly from YAC or cosmid DNA. This strategy was attempted for the isolation of *Not*I sites from YACs 129B7 and 130G11, however this proved unsuccessful.

There are numerous strategies available to identify genes within genomic clones. These methodologies include exon trapping, direct screening of cDNA or CpG island libraries, cDNA selection and direct 'shotgun' sequencing (reviewed by Parrish and Nelson, 1993).

# CHAPTER 4

# MUTATIONAL ANALYSIS OF *Gli2* AND *Inhibinβb* IN *Dh* MICE

### 4.1.1 Introduction

Two genes, *Gli2* and *Inhβb*, have been identified which represent plausible candidates for the *Dh* mutation. Critically, as described in chapter 3, both genes are demonstrated to lie within the genetically defined region of *Dh*. In addition the developmental expression profile and putative functional role of each gene can be considered consistent with a possible role in the *Dh* phenotype. The work presented in this chapter describes analysis for the presence of mutation within the *Dh* allele of these genes.

### 4.1.2 Candidacy of *Gli2* for the *Dh* mutation

The *Gli* family of genes has been implicated in both neoplastic and developmental disorders. Human *GLI* was originally identified by virtue of an approximate 75 fold amplification in a malignant glioma (Kinzler *et al.*, 1987). Subsequently *GLI* was found to be amplified in other glioblastomas (Wong *et al.*, 1987) and in a subset of childhood sarcomas (Roberts *et al.*, 1989), however the role that *GLI* plays in the neoplastic process remains unclear.

Cloning of *GLI* (Kinzler *et al.*, 1988) revealed the presence of five tandemly repeated zinc finger motifs, of the $C_2$-$H_2$ subclass, related to those of the *Xenopus* transcription factor IIIA (TFIIIA), (Miller *et al.*, 1985) and the *Drosophila* gap gene *krüppel* (Rosenberg *et al.*, 1986). On the basis of sequence homology in the zinc finger region the *Gli* family in humans and mice has been extended to three members, *Gli*, *Gli2* and *Gli*3 (Ruppert *et al.*, 1988; Ruppert *et al.*, 1990; Hui *et al.*, 1994; Hughes *et al.*, 1997). Work presented in this thesis, chapter 5, strongly suggests that the mammalian *Gli* family is limited to these three members. In addition two

invertebrate *Gli* genes have been isolated; the *Drosophila* segment polarity gene *cubitis interruptus*, *ci*, (Orenic *et al.*, 1990) and the *Caenorhabditis elegans* sex determining gene *tra1* (Zarkower and Hodgkin, 1992). Comparison of the primary sequence and genomic organisation of the *Gli* zinc finger regions with related zinc finger genes indicate that they constitute a distinct family of zinc finger genes (Ruppert *et al.*, 1988). A characteristic of the *Gli* zinc finger consensus sequence, $(Y/F)XCX_3GCX_3(F/Y)X_5LX_2HX_{3-4}H(T/S)GEKP$, is however the H-C link, H(T/S)GEKP(Y/F)XC, which connects the histidine of one finger to the cysteine of the next. This sequence is a defining feature of the eukaryotic krüppel-related zinc finger super family (Schuh *et al.*, 1986; Chowdhury *et al.*, 1987). By using an oligonucleotide probe degenerate to the H-C link, Bellefroid *et al.* (1989) estimate the human genome to contain approximately 300 krüppel-related genes.

In addition to the homologous zinc finger domains the mammalian *Gli* genes posses additional regions of homology. Alignment of *GLI* and *GLI3* identifies seven collinear regions of homology (Ruppert *et al.*, 1990); when mouse *Gli2* and human *GLI3* are compared these regions are conserved, extended and an additional six regions of homology are detected (Hughes *et al.*, 1997). Presumably these domains perform either a structural function or are involved in protein, protein interactions.

The GLI proteins are proposed to be transcriptional regulators based on DNA binding activity and the demonstration that GLI protein localises to the nucleus (Kinzler and Vogelstein, 1990; Ruppert *et al.*, 1990; Vortkamp *et al.*, 1995). The high level of homology within the zinc fingers, particularly fingers 3 - 5, predicts that the GLI proteins could interact with the same or very similar target DNA sequences. Indeed it has been shown that GLI, GLI3 and TRA1 bind very similar sequences *in vitro* (Kinzler and Vogelstein, 1990; Ruppert *et al.*, 1990; Vortkamp *et al.*, 1995; Zarkower and Hodgkin, 1993). Crystallographic studies by Pavletich and Pabo (1995) have demonstrated that fingers 4 and 5 are responsible for the main base contacts to a high affinity 9 bp consensus sequence. Identification of consensus DNA target sequences should aid in the identification of the *in vivo* targets of the

GLI proteins. Indeed, in a recent report by Alexandre *et al.* (1996), consensus GLI binding sites are identified in the promoter region of the *Drosophila patched* (*ptc*) gene which are shown to be necessary for the induction of *ptc* expression by *hedgehog*. Also Sasaki *et al.* (1997) demonstrate that a GLI binding site is required for the activity of the enhancer for floor plate expression of Hepatocyte Nuclear Factor-3β, *HNF-3β*, *in vivo*. In addition they demonstrate that, *in vitro*, GLI but not GLI3 is able to activate this binding site.

The three mouse *Gli* genes are widely expressed in ectoderm and mesoderm derived tissues during post implantation development (Hui *et al.*, 1994; Mo *et al.*, 1997; Walterhouse *et al.*, 1993; Hui and Joyner 1993). *Gli2* and *Gli*3 show an almost identical expression pattern throughout development, while the pattern for *Gli*, although displaying significant overlap is distinct. *Gli2* expression is first detected during gastrulation (day 7.5) throughout the mesoderm and ectoderm but is not present in the endoderm of the embryo. At neurulation stages (day 8.5) *Gli2* is expressed in the dorsal region of the neural tube with highest levels in forebrain and midbrain areas. In mesoderm, expression was detected in the somites, presomitic mesoderm, cranial mesenchyme of neural crest origin and in the somatic and splanchnic components of the lateral mesoderm.

During organogenesis (days 10.5-17.5) the *Gli* genes are widely expressed in ectoderm and mesoderm derived structures. In the developing brain and spinal cord the regions where the *Gli* genes are expressed correlate with the ventricular zones suggesting expression in mitotically active cells or cells undergoing differentiation. As in the neural tube *Gli2* expression is restricted dorsally. Other sites of ectodermal *Gli2* expression include the optic stalk and lens while low levels are detected in Rathkes' pouch, from which the anterior pituitary develops.

Expression of the *Gli* genes is also detected in non neuronal derivatives of cranial neural crest, which contribute to most of the skeletal and connective tissues of the head. Expression is however not detected in the neuronal derivatives of cranial neural crest. At day 9.5 *Gli2* is expressed in crest derived facial and pharyngeal arch

mesenchyme and from day 10.5 to 12.5 *Gli2* transcripts are found to be uniformly distributed in most craniofacial structures. At day 14.5 *Gli2* is found in the cartilage forming cells of the otic capsule, the optic capsule, nasal capsule and the base of the skull. At day 14.5 the tooth buds and mesenchyme surrounding the whisker follicles are also sites of *Gli2* expression.

From day 11.5 to 14.5 *Gli2* is expressed in the following mesoderm derived structures; the mesenchyme of the gut (derived from the splanchnic component of the lateral mesoderm), the body wall of the thoracic and abdominal regions (derived from the somitic component of the lateral mesoderm), the prevertebrae and in the mesenchyme of the developing lung, mesonephros and testes. In both fore and hind limb buds, at day 10.5, *Gli2* is broadly expressed in the undifferentiated mesenchyme, however the region of posterior mesenchyme which represents the ZPA and expresses *Sonic hedgehog*, *Shh*, does not express *Gli2* (or both *Gli* and *Gli*3). At day 11.5 *Gli2* expression becomes localised to the mesenchyme surrounding the blastema (condensing mesenchyme) and by day 12.5 becomes further restricted to the interdigital mesenchyme.

The extensive expression pattern of *Gli2* suggests it plays multiple roles during postimplantation development. This has recently been substantiated by the generation of *Gli2* null mutant mice through the technology of gene targeting (Mo *et al.*, 1997). Animals heterozygous for the *Gli2* mutation display no detectable abnormalities but homozygotes display a range of skeletal defects. These include cleft palate, tooth abnormalities, absence of vertebral body and intervertebral discs and a reduced sternum. Limbs were significantly shortened and ossification of the long bones and digits was severely delayed. Visceral abnormalities have not been reported.

A role for *Gli*3 during development has also been demonstrated by molecular analysis of the human disorders Greig cephalopolysyndactyly syndrome (GCPS) and Pallister-Hall syndrome (PHS) and the mouse mutant *Extra toes* (*Xt*), which

shows many similarities to the *Dh* phenotype (see chapter 1, 1.3.5, for a description of the *Xt* and GCPS and PHS phenotypes).

Analysis of balanced translocations and deletions in GCPS patients defined the GCPS locus to be on 7p13, (Tommerup and Neilsen, 1983; Kruger *et al*, 1989; Wagner *et al*, 1990). Analysis of genes known to map to 7p13 identified *GLI3* as being disrupted by balanced translocations in three GCPS patients (Vortkamp *et al*, 1991). In two cases the translocation breakpoint occurred within the first third of the coding region of *GLI3* and in the third case the breakpoint was located 10kb downstream of the 3' end of *GLI3*. Recently Wild *et al.* (1997) report point mutations in *GLI3* in two GCPS patients. In one case a nonsense mutation generates a stop codon truncating the protein in the C-H link of the first zinc finger. In a second case a missense mutation causes a Pro→Ser transition, altering a potential phosphorylation site.

Investigation of two *Xt* alleles, *Xt* and $Xt^{J}$ identified non-overlapping deletions in the 5' and 3' of *Gli*3 respectively. Both deletions cause an absence of *Gli*3 expression (Vortkamp *et al*, 1992; Schimmang *et al*, 1992; Hui and Joyner, 1993). This evidence strongly suggests that *Xt* is a mouse model for GCPS and that mutations leading to haploinsufficiency of GLI3 are responsible for the *Xt* heterozygous phenotype and for GCPS.

Frameshift mutations in *GLI3*, caused by single base pair deletions, were demonstrated in two PHS families (Kang *et al.*, 1997). In both cases this would result in a protein truncated just 3′ of the zinc finger domain. The mode of action of such a truncated *GLI3* protein is unknown but due to the more severe phenotype seen in PHS compared to GCPS it is unlikely that the PHS alleles are simply causing haploinsufficiency. The PHS alleles may be acting as antimorphs, in a dominant negative fashion. Alternatively the truncated protein may be acting as a neomorph by binding to additional promoter regions compared to the wild type *GLI3* thereby perturbing developmental pathways outside of the normal *GLI3* pathway.

Two further *Xt* alleles have been described. Pohl *et al.* (1990) described $xt^{add}$ (anterior digit deformity), a recessive allele caused by a transgene insertion 44kb upstream of *Gli3* which causes a reduction in *Gli3* expression (van der Hoeven *et al*, 1993). Schimmang *et al.* (1994) documented the $Xt^{pdn}$ (Polydactyly Nagoya) allele in which *Gli3* is neither deleted nor is there any diminution in *Gli3* expression, however the *Gli3* coding sequence has not been analysed for mutation.

Comparison of the *Gli2* expression pattern with the tissues which are affected by *Dh* shows that every structure which displays a phenotype expresses *Gli2* during development. The earliest defect to be observed in *Dh* occurs in the splanchnic component of the lateral mesoderm (Green, 1967) a region of high *Gli2* expression. A diminished or absent epithelial structure in the posterior region of the splanchnic mesoderm is postulated to be causal in the failure of development of the spleen. Also high levels of *Gli2* expression in the gut mesenchyme and mesonephros correlate with the phenotypic features of the alimentary canal and kidneys. A similar correlation can be made for the skeletal defects observed in *Dh*; *Gli2* is expressed in the developing limbs, vertebrae, ribs and sternebrae.

There are many areas of *Gli2* expression where no overt *Dh* phenotype is seen. Only the posterior half of the animal is affected in *Dh*; there is no craniofacial, neurological or forelimb phenotype. This apparent discrepancy is however well documented; mutation very often does not effect every tissue in which the mutated gene is normally expressed. Also *Dh* is known to be a gain of function mutation (see chapter 1, 1.3.3). This may be due to a mutant GLI2 protein which is able to aberrantly interact and interfere with specific cellular processes in only a subset of *Gli2* expressing tissues. Alternatively the mutation may affect the normal temporal and/or spatial pattern of *Gli2* mRNA or GLI2 protein expression in only those tissues affected in *Dh*. A further consideration is that *GLI2* and *GLI3* are highly homologous proteins and display almost identical expression patterns. It is therefore possible that, in a subset of tissues, *Gli3* is able to compensate for a mutation in *Gli2* and therefore prevent a phenotype.

The candidacy of *Gli2* for *Dh* is further enhanced by the position of the *Gli* genes within the vertebrate *hedgehog* signalling pathways. As discussed in chapter 1, (1.2.3) *Sonic hedgehog* (*Shh*) plays a critical role in pattern formation of embryonic tissues including the limb, while *Indian hedgehog* (*Ihh*) is involved in the regulation of chondrogenic development. Direct involvement of the *Gli* genes with the *Shh* signal has been reported by Marigo *et al.* (1996) who show *Gli* and *Gli3* to be differentially regulated by *Shh* during limb development and that *GLI* can induce expression of *Patched* (*Ptc*), a known target of *Shh*. Further, ectopic expression of *Shh* in the anterior limb bud induces digit duplication reminiscent of the polydactyly seen in the luxoid mouse mutants. Indeed ectopic *Shh* expression in the anterior limb bud has been reported in *Xt, Hx, Lx* and *Lst* (Chan *et al.*, 1995; Masuya *et al.*, 1995; Masuya *et al.*, 1997). *Dh* may therefore result from a disruption of normal hedgehog signalling.

Taken together, the genetic location of *Gli2* within the *Dh* critical region, the developmental expression pattern of *Gli2* in those structures affected by *Dh*, the functional role demonstrated for the *Gli* genes during development and the position of the *Gli* genes within the hedgehog signalling pathway identify *Gli2* as a strong candidate for *Dh*.

### 4.1.3 Candidacy of *Inβb* for the *Dh* mutation

Inhibins and activins are members of the transforming growth factor β (TGFβ) family of signalling molecules. They were originally isolated from mammalian follicular fluid on the basis of their ability to respectively inhibit or stimulate the release of follicle stimulating hormone (FSH) from the anterior pituitary (reviewed by Vale *et al.*, 1990). In addition to their action on FSH release, their ability to regulate secretion of other hypothalamic or pituitary peptides such as oxytocin, adrenocorticotrophic hormone (ACTH) and growth hormone (GH) has been shown (Vale *et al.*, 1990). Inhibins and activins are dimers assembled from subunits encoded by three distinct but related genes; *Inhβa*, *Inhβb* and *Inhα*. The

precursor polypeptides are proteolytically processed into mature peptides which assemble into dimers to form the biologically active molecules. Inhibins are α:βA or α:βB heterodimers, whereas activins are βA:βA or βB:βB homodimers or βA:βB heterodimers. Recently additional members of the inhibin/activin β subunit family have been isolated, *Inhβc*, *Inhβd* and *Inhβe* which are more closely related to each other than they are to *Inhβa* or *Inhβb* (Hotten *et al.*, 1995; Lau *et al.*, 1996; Fang *et al.*, 1996; Oda *et al.*, 1995). The function of *Inhβc*, *Inhβd* and *Inhβe* is not yet known, although in the adult mouse *Inhβc* and *Inhβe* exhibit a liver specific expression pattern (Fang *et al.*, 1997) and *Inhβd* can induce a secondary body axis and mesoderm induction in Xenopus embryos (Oda *et al.*, 1995). It is currently unclear if *Inhβc Inhβd* and *Inhβe* can assemble into homodimers, heterodimers, or both.

Activins exert their biological functions through binding to specific serine/threonine kinase receptors. As reviewed by Matthews (1994), the two known type II activin receptors, ActR-II and IIB, are constitutively active kinases which when bound to activin can associate and phosphorylate type I receptors, four of which are currently described; ActR-I, ActR-IB, ALK-2 (activin like receptor) and ALK-4. These heterodimeric receptor/activin complexes then go on to initiate intracellular signalling. In addition to these cell surface receptors activin activity is modulated by the follistatins. Six isoforms of follistatin have been identified which are generated by C-terminal truncation and/or glycosylation (for review see Michel *et al.*, 1993). Follistatins are able to bind to and inhibit the activity of activin in a number of systems (see Hashimoto *et al.*, 1992 and refs. therein). However additional defects in follistatin deficient mice (Matzuk *et al.*, 1995a) compared to mice lacking activin (Matzuk *et al.*, 1995b) suggest additional roles for follistatin which are independent of activin. The regulation by follistatin of signalling by other TGFβ-related molecules is suggested (Matzuk *et al.*, 1995a).

Activins are proposed to play multiple roles in early development. High levels of activin are found in the oocyte and fertilised egg (Albano *et al.*, 1993) consistent with the observation that activin A is a potent stimulator of oocyte maturation (Itoh *et al.*, 1990). During blastocyst formation activin displays a dynamic distribution between cells of the inner cell mass and the trophectoderm suggesting a role in the modulation of growth or the inhibition of differentiation (Albano *et al.*, 1993). During early postimplantation development *Inhα* transcripts are not detected while *Inhβa* and *Inhβb* transcripts are restricted to the deciduum, their function therein remaining unclear (Albano *et al.*, 1994). From day 10.5 - 12.5 activin subunits are reexpressed in the many areas of the embryo (Feijen *et al.*, 1994). *Inhβb* is strongly expressed in selected regions of the brain, particularly in areas of the fore and hindbrain and spinal cord which are undergoing rapid cellular proliferation. Expression is also seen in the epithelial lining of the oesophagus and glandular part of the stomach, the mesenchymal condensations of the tarsus in the developing eyelid, the mesenchyme in the wall of the large blood vessels in the heart region, the gonad primordium, the pleuroperitoneal membrane and the intervertebral disc anlagen.

Feijen *et al.* (1994) also report *Inhβb* expression in the developing limb, although they do not specify a detailed pattern of expression. *Inhβa* is also shown to be expressed in the limb at day 10.5 - 12.5 and is further detailed to be restricted to precartilage condensations. Support for the involvement of activin in limb morphogenesis comes from the treatment of limb bud mesenchyme which can induce chondrogenesis (Chen *et al.*, 1993; Jiang *et al.*, 1993). Also ectopic application of TGFβ-1, which signals through receptors structurally related to the ActR-IIs (Lin *et al.*, 1992) causes alterations in the skeletal elements of the developing chick limb in a stage and position specific manner (Hayamizu *et al.*, 1991).

The pattern of expression for *Inhβb* has many inconsistencies with those tissues which are affected by *Dh*, for example no *Inhβb* expression is detected in the

splanchnic component of the lateral mesoderm or the developing kidney, while high levels are seen in the CNS, an area which displays no phenotype in *Dh*. However this does not preclude mutation in *Inhβb* from being responsible for *Dh*, as mutations can be unpredictable in their sites of action, particularly mutations causing mis-expression. Indeed *Inhβb* null mice were predicted to display a severe developmental phenotype, including disrupted mesoderm induction, however the only defect observed is a failure of eyelid fusion leading to eye lesions (Vassalli *et al.*, 1994). Further the nature of activin as an inductive signalling molecule, active at many stages of development, not least in the developing limb, implicates *Inhβb* as a plausible candidate for *Dh*.

## 4.2 Results

### 4.2.1 Analysis of the *Gli2* coding sequence in *Dh* mice

Southern blot analysis, performed by David Hughes, failed to detect any obvious deletions or other rearrangements in *Dh* mice relative to wild type. A systematic strategy was therefore undertaken to identify any possible mutation within the coding region of the *Gli2* transcript. This was approached by RT-PCR amplification of the entire coding sequence of *Gli2* from E12.5 homozygote *Dh* RNA in overlapping segments of approximately 1kb. This resource could then be analysed for the identification of coding sequence and splice site and mutations.

To isolate *Dh* homozygote RNA *Dh* heterozygotes, maintained on a C57BL/6 background, were mated and embryos taken at 12.5 days of gestation. Total RNA was then isolated from each embryo in the litter (this procedure was performed by D. Hughes). Embryos of age 12.5 days were chosen as *Gli2* has been shown to be expressed at high levels in a range of tissues at this age (Hui *et al.*, 1994). The genotype of each embryo was then established using a restriction site polymorphism previously detected in the 5′ coding region of *Gli2* by D. Hughes; RT-PCR using the primer pair 12a/1b amplifies a 279 bp fragment encompassing nucleotide 932 (based on the *Gli2* cDNA sequence, Hughes *et al.*, 1997) which is G in *Dh* and T in C57BL/6. This polymorphism creates a *Hpa*II site in *Dh* which is absent in C57BL/6. The genotype of each animal in the litter is shown in figure 4.1. Two embryos were identified as *Dh* homozygotes, six as heterozygotes and one as wild type.

PCR primers were designed, based on the *Gli2* cDNA sequence, such that the entire coding sequence could be RT-PCR amplified in five overlapping fragments. The primers and fragments amplified are listed in table 4.1. Primers 5a/5b did not successfully amplify the 3′ most fragment. Additional primers were therefore designed to facilitate the amplification of smaller fragments, see table 4.1. Amplification of all fragments was successful only with the addition of the PCR

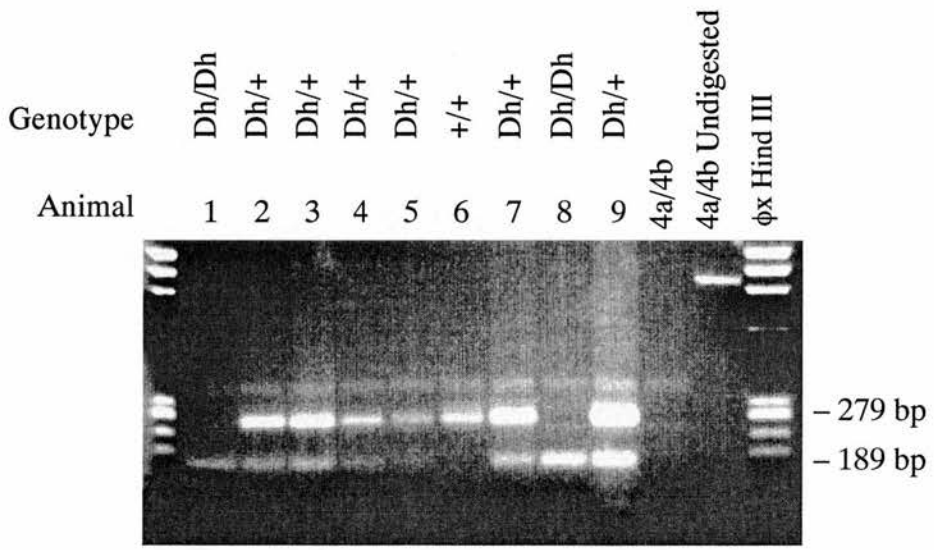



**Figure 4.1 Genotype of embryos derived from a *Dh*/+ × *Dh*/+ mating**

RT-PCR amplification from nine (*Dh*/+ × *Dh*/+) embryos with primers 12a/1b followed by *Hpa*II digestion. The 279bp wild type allele (C57BL/6) is digested to fragments of 189bp and 90bp, while the *Dh* allele contains no *Hpa*II site. A 950bp *Gli2* fragment, 4a/4b, was included in each digestion to control for complete digestion of the 12a/1b fragment. For each animal 4a/4b is completely digested indicating complete digestion of the wild type 12a/1b fragment. The genotype of each animal is shown.

additive, Taq Extender (Stratagene) which contains the thermostable polymerase *Pfu*. This enzyme has a 3′ - 5′ proof-reading activity which reduces premature chain termination due to the incorporation of mis-matched base pairs. This enhances efficiency of template extension by *Taq* polymerase, resulting in a greater yield of product. Importantly, for mutation analysis, it also enhances amplification fidelity (Cline *et al.*, 1996).

**Table 4.1 Primers used for the RT-PCR amplification of the *Gli2* coding sequence from *Dh* mice**

Primers are listed 5′ - 3′. Nucleotide position is according to the *Gli2* cDNA sequence (Hughes *et al.*, 1997), where the initiation codon is nucleotide 110 and the termination codon is 4744

| Fragment and position of primers | Forward primer | Reverse primer |
|---|---|---|
| 1a/1b; 80 - 1129 | GTCCGCCACCAAAGAGTTTGAGCC | AGTTGGGTAGGCATGGTGCTGA |
| 2a/2b; 1040 - 2065 | CCCCTGATCCAGCCTTCACCCAC | AGAGGGCTCGCTGCTGCAGGATG |
| 3a/3b; 2006 - 3033 | AGCTCCGGGCTTTGTCAGTCCAG | CTGTGGAAACGTTGCACTCGAGGC |
| 4a/4b; 2947 - 3892 | GGGGCCAAACAGCAGCACACGG | GCTGGGCTTGGCTGCCATATTGC |
| 5a/5b; 3817 - 4781 | GCTGAGCCCAAACATTGTCAGCGG | GTGTTGCTCAGGTCTCCTGTGCCA |
| 5a/5d; 3817 - 4141 | GCTGAGCCCAAACATTGTCAGCGG | GGGTTCCAGGGGTGGCCGAGG |
| 5c/5e; 4077 - 4458 | GCTTCATGGAGTCCCAGCAGAAC | CCCGGTGAAGGCAAAGGCTCAG |
| 5f/5b; 4355 - 4781 | CACATGTATGAACAGAATGGAG | GTGTTGCTCAGGTCTCCTGTGCCA |

Figure 4.2 shows the amplification products for all primer pairs from *Dh* homozygote, heterozygote and wild type RNA. For each segment, amplification from wild type RNA produced one band of the expected size only. This indicates the absence of alternatively spliced transcripts which is consistent with the observation of only one *Gli2* transcript of 8kb seen between day 8.5 to 16.5 of development (Hui *et al.*, 1994). (In addition to the expected band of 324bp, a second band of 420bp is seen amplified by primers 5a/5d but this band was not reproduced in subsequent experiments and was therefore considered artifactual).

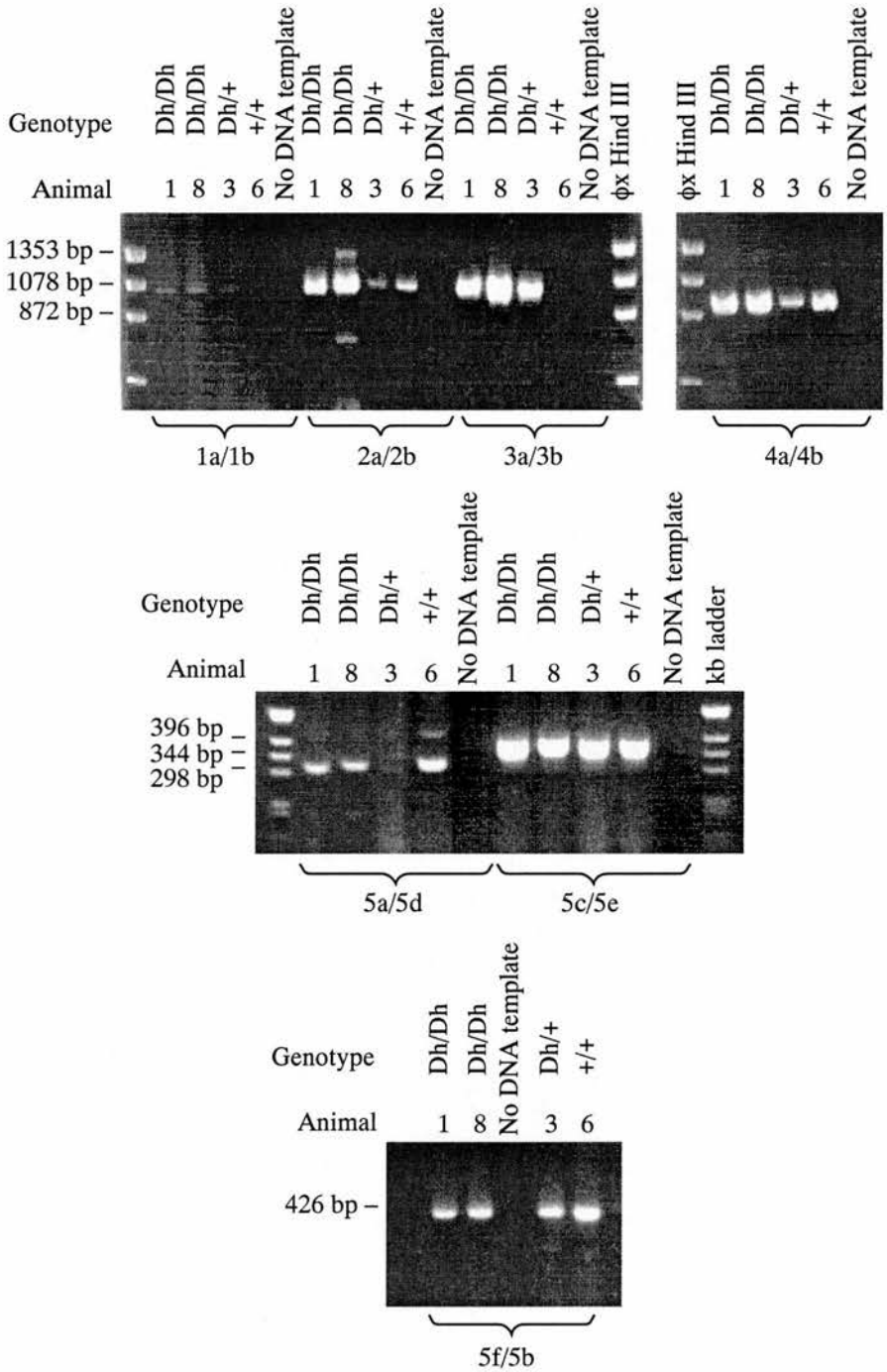



**Figure 4.2 RT-PCR amplification of *Gli2* from wild type and *Dh* animals**
RT-PCR amplification of overlapping fragments which cover the *Gli2* coding region from +/+, *Dh*/+, *Dh*/*Dh* embryos. For each fragment only one band of the expected size is consistently amplified from each genotype. The identity of each fragment is indicated below each panel.

Amplification of all fragments from *Dh* homozygotes and heterozygotes also consistently produced bands of the expected size only. Firstly, this confirms that *Gli2* is expressed in *Dh* animals. Secondly, this demonstrates the absence of a mutation which grossly affects the size of the coding region of the *Gli2* transcript.

To further analyse the *Gli2* coding region, the RT-PCR products from the *Dh* homozygotes were cloned to enable each fragment to be unambiguously sequenced. To ensure that the sequence analysis was performed on DNA derived from the *Dh* chromosome every attention was paid during the RT-PCR procedure to eliminate possible contamination of *Dh* homozygous reactions with wild type or *Dh* heterozygote RNA or cDNA. Homozygous reactions were set up separately and the control reaction containing no template cDNA was ensured to be free from product. Only fragment 1a/1b was amenable to typing based on the *Hpa*II restriction site polymorphism at nucleotide 932. To control for contamination during amplification of the other fragments primer pair 12a/1b was used to amplify from the same reverse transcription reactions from which each of the other fragments were amplified. Digestion of these fragments with *Hpa*II showed that there was no cross contamination.

RT-PCR product was cloned directly into the vector pCRII (Invitrogen) and at least three positive clones identified by restriction analysis of DNA isolated from colonies selected on the basis of α-complementation. Double stranded DNA from clones containing each fragment was then prepared and sequenced from both strands using the vector primers M13 forward and M13 reverse and primers designed internal to each fragment. Fragment 3a/3b was sequenced by Gordon Allan and fragment 4a/4b by Laura Lettice. Each fragment was analysed until unequivocal sequence was obtained and compared to the known *Gli2* cDNA sequence. Sequence variation was initially investigated by sequence analysis of DNA isolated from independent clones of the RT-PCR products. This served to identify variation due to PCR infidelity. A total of seven single base pair substitutions were detected, presumably caused by mis-match incorporation.

Determination of the entire coding sequence of *Gli2* from *Dh* failed to detect a mutation. However a total of four sequence variations were demonstrated in the *Gli2* coding sequence of *Dh* compared to C57BL/6 which are listed in table 4.2. All changes represent single base pair substitutions. Two of these changes are silent and cannot therefore represent a mutation; nucleotide 932 is T in wild type and substituted for C in *Dh* while nucleotide 976 is also T in wild type and substituted for C in *Dh*.

The two other sequence variations detected are predicted to result in single amino acid substitutions in *Dh* compared to C57BL/6 and could therefore represent the *Dh* mutation. Nucleotide 241 is C in C57BL/6 and A in *Dh* which causes the substitution at amino acid 44 of histidine → glutamine. This is a non-conservative change; histidine has a basic side chain ($CH_2CNHCHNHCH$) while glutamine has an amide side chain ($CH_2CH_2CONH_2$) and changes an amino acid which is conserved in *Gli3* but not *Gli*. In order to determine whether the histidine$^{44}$ → glutamine substitution represents the *Dh* mutation other inbred strains were analysed for the presence of this change. The *Dh* mutation arose in a male mouse derived from either CBA, C3H or 101. Sequence analysis of RT-PCR product encompassing nucleotide 241 from each of these inbred strains demonstrated the same sequence as the *Dh* allele. Any variant present in any other strains cannot be the basis for the *Dh* mutation. The histidine$^{44}$ → glutamine difference is therefore a polymorphism and not the *Dh* mutation.

Another base pair substitution was observed at nucleotide 446; G in wild type is substituted for A in *Dh* resulting in the substitution of alanine$^{113}$ → threonine. This is a semi-conservative change; alanine has an aliphatic side chain ($CH_3$) whereas threonine has an aliphatic hydroxyl side chain ($CHOH\text{-}CH_3$) and changes an amino acid which is not conserved in *Gli* or *Gli3*. To assess this candidate the occurrence of threonine$^{113}$ was determined in a range of inbred strains. The substitution of G$^{336}$ → A fortuitously disrupts a *Hha*I restriction site, therefore PCR

primers, I51/I52, flanking this site which would amplify from genomic DNA were designed. The amplified fragments from nine inbred strains were then subject to digestion with *Hha*I, figure 4.3. Six of the stains tested failed to cut with *Hha*I suggesting a G in position 336. This was confirmed by sequence analysis of the I51/I52 PCR product from C3H and DBA which demonstrated the same sequence as *Dh*. The alanine$^{113}$ → threonine polymorphism is therefore not responsible for the *Dh* phenotype.

**Table 4.2 Summary of sequence variation detected in *Gli2***

Nucleotide sequence variation detected between C57BL/6 and *Dh* alleles of *Gli2* and corresponding amino acid changes where applicable. Each sequence variation causing an amino acid substitution was also detected in a number of inbred strains (see text).

| Nucleotide position | C57BL/6 | *Dh* | Amino acid position | C57BL/6 | *Dh* |
|---|---|---|---|---|---|
| 241 | C | A | 44 | His | Gln |
| 446 | G | A | 113 | Ala | Thr |
| 932 | T | C | Silent | | |
| 976 | T | C | Silent | | |

### 4.2.2 Expression analysis of *Gli2* in *Dh* mice

Semi-quantitative RT-PCR analysis, performed by David Hughes, had previously indicated that the expression of *Gli2* in 10.5 day *Dh* animals was unaffected compared to wild type. However in the course of genotyping a litter of embryos from a *Dh*/+ × *Dh*/+ mating an apparent under-expression of *Gli2* was observed in *Dh* heterozygotes. Figure 4.1 shows the litter typed on the basis of the polymorphism at nucleotide 932. In heterozygotes, animals 2, 3, 4, 7 and 9, the *Dh* and wild type alleles should be amplified at the same rate but the wild type allele, 279bp, is judged to be at an approximate 3-4 fold intensity compared to the *Dh* allele

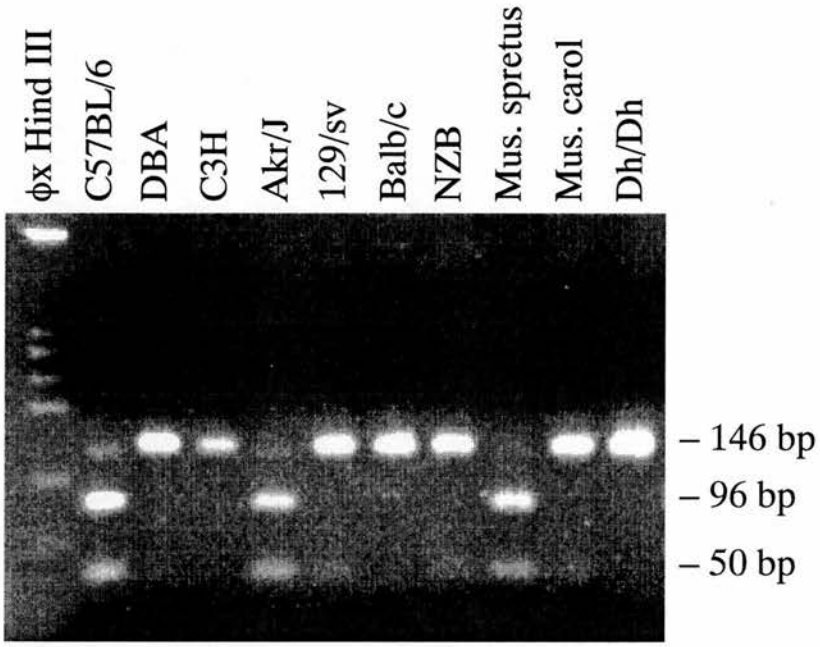



**Figure 4.3 The occurrence of *Gli2* $A^{336}$ allele**

I51/I52 PCR amplification from a selection of inbred strains followed by *Hha*I restriction digestion and resolution on a 5% Nusieve agarose gel. The *Hha*I site present in C57BL/6 results in fragments of 96bp and 50bp but $A^{336}$ in *Dh* disrupts this site and the 146bp fragment remains undigested. The fragment from strains DBA, C3H, 129/Sv, Balb/c, NZB and *Mus caroli* also remains undigested suggesting the $G^{336}$ substitution is not unique to *Dh*.

of 189bp. This is unlikely to be accounted for by the difference in band size or an artefactual effect of electrophoresis and could therefore indicate a genuine reduction of *Gli2* expression in *Dh* mice.

To investigate this possibility RNA from the same litter was analysed by RT-PCR using the restriction site polymorphisms at nucleotides 446 and 932. The PCR was performed as standard except for the incorporation of $P^{32}$ end labelled primers I51 and 1b for amplification using I51/I52 and 12a/1b respectively. This enables quantification of alleles by phosphorimage technology. For both assays the undigested band seen in heterozygotes is at least twice as intense as the band produced following digestion (figure 4.4). Hence in assay 12a/1b - *Hpa*II, *Gli2* appears to be under-expressed in *Dh*, as in figure 4.1, while in assay I51/I52 - *Hha*I, *Gli2* appears to be over-expressed in *Dh*. This cannot therefore represent a genuine difference in expression level. A plausible explanation is that during the final rounds of PCR, when the reaction has reached saturation, single strand DNA reanneals to complementary strands due to the reduction in DNA synthesis. This results in heteroduplex fragments, a *Dh* strand annealed to a wild type strand which will disrupt the polymorphic *Hpa*II or *Hha*I sites. Consequently the undigested allele in each assay is disproportionately high.

### 4.2.3 Cloning and sequencing of the mouse *Inhβb* gene

At the time of this study only a portion of the mouse *Inhβb* gene sequence was available (Ritous *et al.*, 1995). Comparison of this sequence with that of the rat (Feng *et al.*, 1989; Esch *et al.*, 1987) and human (Mason *et al.*, 1989) indicated that it covered the majority of the coding sequence in exon two (the rat and human *Inhβb* genes consist of two exons). Seven coding nucleotides were predicted to be missing at the 5′ end of the exon and 63 at the 3′ end. To perform mutational analysis it was therefore necessary to determine the entire coding sequence, such that the *Dh* allele could be compared to wild type.

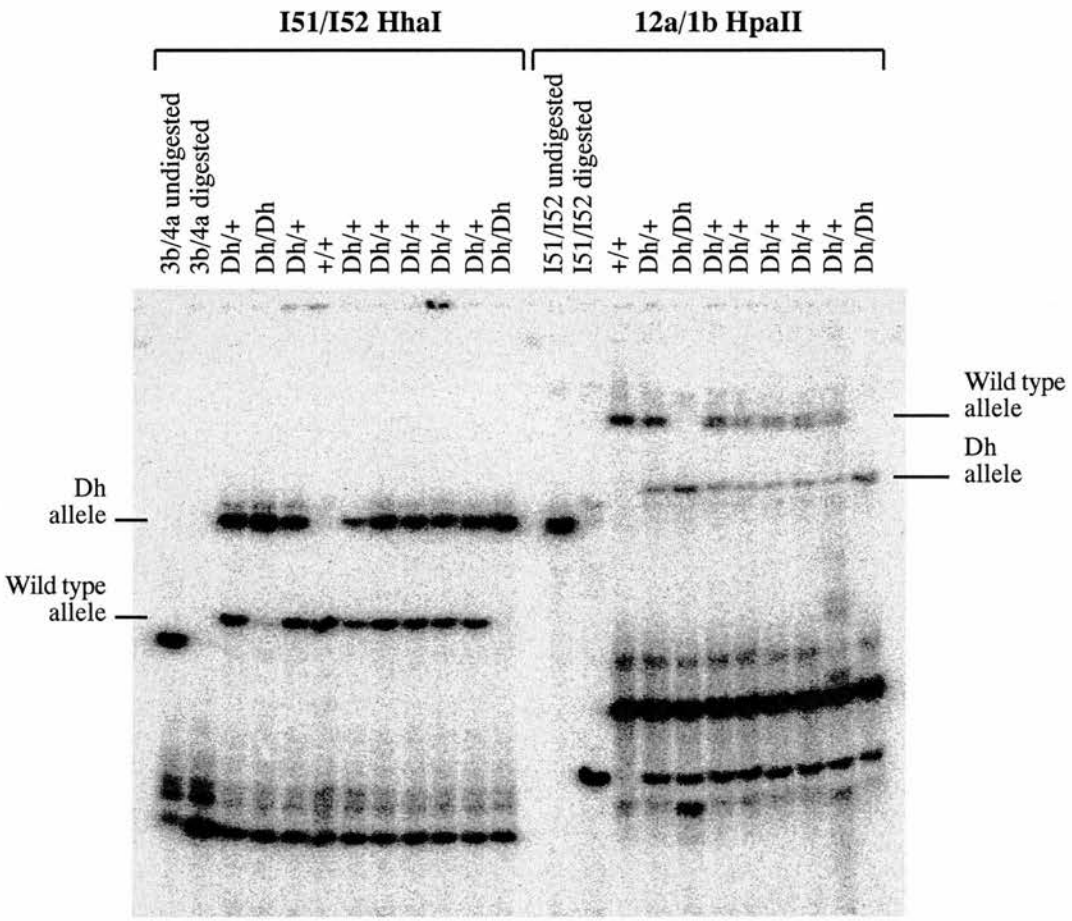



**Figure 4.4 Relative expression of *Gli2* from *Dh* and wild type alleles in 12.5 day *Dh* mice**

The *Gli2* fragments I51/I52 and 12a/1b were amplified by RT-PCR from a day 12.5 *Dh*/+ × *Dh*/+ litter. The PCR was performed as standard except for the incorporation of $P^{32}$ end labelled primers I51 and 1b. Amplified product was then digested with either *Hha*I or *Hpa*II, as shown. A fragment containing the appropriate restriction site was added to each sample to control for complete digestion; 3b/4a in I51/I52 - *Hha*I and I51/I52 in 12a/1b - *Hpa*II. Following digestion the products were resolved on a 6% non-denaturing polyacrylamide gel, the gel dried and analysed by phosphorimagery.

Isolation of a lambda clone containing *Inhβb* was previously described in chapter 3 (3.2.1). To subclone genic fragments amenable to sequencing a vectorette-PCR strategy was initially employed. The lambda clone was digested to completion with either *BamH*I, *Bcl*I, *Bgl*II or *Sau*3AI and a vectorette possessing the compatible overhang, 5′ GATC 3′, ligated onto the digestion products. These vectorette libraries were then amenable to PCR performed using the vectorette primer, G400, and one of a set of *Inhβb* specific primers derived from the mouse *Inhβb* sequence or, for exon one, from rat and human sequence. Where the *Inhβb* specific primers are located a suitable distance from the vectorette then, under appropriate reaction conditions, amplification of the intervening sequence should be achieved.

Despite numerous modifications to the PCR reaction parameters only one genic fragment was successfully produced. Amplification from the *Sau*3A1 vectorette library using G400 and I773, located in the 3′ region of exon two (see table 4.3) enabled the amplification of a 106 bp fragment. Sequence analysis then showed this fragment not to consist of the entire unknown 3′ coding sequence due to a *Sau*3A1 site located 24 bp upstream of the termination codon.

A second strategy to subclone the *Inhβb* gene was therefore undertaken based on subcloning the lambda clone into pBS. The lambda insert was isolated from the lambda arms by electrophoresis in 0.6% agarose following *Sal*I digestion. Insert DNA was then digested to completion with *Sau*3AI and shotgun cloned into the *BamH*I site of pBS. Approximately 300 resultant colonies, which represents an estimated six fold redundancy, were then screened by hybridisation using five different $P^{32}$ end labelled oligonucleotides, table 4.3. Oligonucleotide I181, derived from the available mouse sequence, is complementary to 5′ sequence of exon two and I942, derived from rat and human sequence, is complementary to sequence 3′ of the *Sau*3AI site identified by vectorette PCR at nucleotide 1109. Three oligonucleotides, I940, I771 and I941, complementary to exon 1 sequence were designed based on the rat and human sequences. Limited degeneracy was introduced into I771.

Two positive colonies were identified for I940, I771 and I942, one colony was positive for I941 while oligo I181 failed to identify a positive clone. Inserts were sequenced using pBS vector primers and *Inhβb* primers such that *Inhβb* coding sequence was determined from both strands. Due to the failure to identify a subclone with oligo I181 the seven most 5′ nucleotides of exon two could not be determined in this way. The entire sequence of exon 2 was therefore subsequently determined by sequence analysis of RT-PCR product amplified from wild type RNA using primers I941 and J96. The complete coding sequence determined for mouse *Inhβb* is illustrated in figure 4.5.

Relative to the published sequence of exon 2 (Ritous *et al.*, 1995) one anomaly was observed. Nucleotide 1023 was determined to be G compared to A. This is a silent change but results in the creation of a *Hpa*II site. The presence of this site was assayed in a number of inbred strains, figure 4.6, and found to be present in every strain analysed. This suggests an error in the sequence of Ritous *et al.*, however the origin of the DNA sequenced by Ritous *et al.* is not stated and could therefore represent a genuine polymorphism if the sequence was derived from a strain not analysed here.

Comparison of the mouse *Inhβb* coding sequence to that of rat and human demonstrates a very high degree of homology. At the nucleotide level mouse *Inhβb* is 97% identical to rat and 90% identical to human. The comparison at the amino acid level is illustrated in figure 4.7. Only two differences are observed between mouse and rat, giving a 99% identity, while there are eleven differences between mouse and human giving a 97% identity. This demonstrates the sequence presented here to indeed represent the mouse *Inhβb* gene.

**Figure 4.5 DNA sequence of the mouse *Inhβb* gene**

The coding sequence, designated in upper case, has been translated into single and three letter code. Non translated sequence is shown in lower case. Nucleotide numbering commences with 1 at the first nucleotide of the initiator methionine (intronic sequence is numbered separately). The sequence -4 to -45, represented by n was not unambiguously determined. The sequence -46 to -74 is highly conserved with that of the rat and is therefore depicted to be located at the same distance from the initiating methionine. The position of the intron was determined from the comparison of sequence from genomic DNA and RT-PCR product. The size of the intron was not determined.

```
   -70                -50                -30
    .        .         .        .         .        .
ggagcccggagcccggccctgagctcgctnnnnnnnnnnnnnnnnnnnnnnnnnnnnn

   -10                 10                 30
    .        .         .        .         .        .
nnnnnnnnnnnaccATGGACGGGCTGCCCGGTCGGGCGTTGGGGGCCGCCTGCCTTCTG
              MetAspGlyLeuProGlyArgAlaLeuGlyAlaAlaCysLeuLeu
               M  D  G  L  P  G  R  A  L  G  A  A  C  L  L

    50                 70                 90
    .        .         .        .         .        .
CTCCTGGTGGCCGGCTGGTTGGGACCCGAGGCCTGGGGCTCTCCCACACCCCCGCCGTCG
LeuLeuValAlaGlyTrpLeuGlyProGluAlaTrpGlySerProThrProProProSer
 L  L  V  A  G  W  L  G  P  E  A  W  G  S  P  T  P  P  P  S

   110                130                150
    .        .         .        .         .        .
CCCGCTGCGCCGCCCCCGCCCCCGCCACCCGGAGCTCCGGGTGGCTCGCAGGACACCTGT
ProAlaAlaProProProProProProProGlyAlaProGlyGlySerGlnAspThrCys
 P  A  A  P  P  P  P  P  P  P  G  A  P  G  G  S  Q  D  T  C

   170                190                210
    .        .         .        .         .        .
ACGTCGTGCGGTGGCGGCGGCGGCGGCTTCCGGCGGCCGGAGGAGCTGGGCCGGGTGGAC
ThrSerCysGlyGlyGlyGlyGlyGlyPheArgArgProGluGluLeuGlyArgValAsp
 T  S  C  G  G  G  G  G  G  F  R  R  P  E  E  L  G  R  V  D

   230                250                270
    .        .         .        .         .        .
GGTGACTTCCTGGAGGCGGTGAAGAGACACATCTTGAGCCGCCTGCAGTTGCGGGGCCGG
GlyAspPheLeuGluAlaValLysArgHisIleLeuSerArgLeuGlnLeuArgGlyArg
 G  D  F  L  E  A  V  K  R  H  I  L  S  R  L  Q  L  R  G  R

   290                310                330
    .        .         .        .         .        .
CCCAACATCACGCACGCTGTCCCCAAGGCGGCCATGGTCACGGCCCTGCGCAAGCTGCAC
ProAsnIleThrHisAlaValProLysAlaAlaMetValThrAlaLeuArgLysLeuHis
 P  N  I  T  H  A  V  P  K  A  A  M  V  T  A  L  R  K  L  H
```

```
   350                 370                 390
    .         .         .         .         .         .
GCCGGCAAGGTGCGCGAGGACGGCCGCGTGGAGATCCCGCACCTCGACGGCCACGCCAGC
AlaGlyLysValArgGluAspGlyArgValGluIleProHisLeuAspGlyHisAlaSer
 A  G  K  V  R  E  D  G  R  V  E  I  P  H  L  D  G  H  A  S

   410                 430                 450
    .         .         .         .         .         .
CCGGGCGCCGACGGCCAGGAGCGCGTCTCCGAGATCATCAGCTTTGCAGAGACAGgtggg
ProGlyAlaAspGlyGlnGluArgValSerGluIleIleSerPheAlaGluThrA
 P  G  A  D  G  Q  E  R  V  S  E  I  I  S  F  A  E  T

  (+20)               (+40)                           470
    .         .         .         .         .   .     .
ttcggccacagggaagccagaggcggcgcgatc--/ Intron /---ATGGCCTCGCC
                                                 spGlyLeuAla
                                                 D  G  L  A

                490                 510                 530
        .         .         .         .         .         .
TCCTCCCGGGTCCGCCTGTACTTCTTCGTCTCTAATGAAGGCAACCAGAACCTATTCGTG
SerSerArgValArgLeuTyrPhePheValSerAsnGluGlyAsnGlnAsnLeuPheVal
 S  S  R  V  R  L  Y  F  F  V  S  N  E  G  N  Q  N  L  F  V

                550                 570                 590
        .         .         .         .         .         .
GTGCAGGCCAGCCTGTGGCTGTACCTGAAACTGCTCCCCTATGTCCTGGAGAAGGGCAGC
ValGlnAlaSerLeuTrpLeuTyrLeuLysLeuLeuProTyrValLeuGluLysGlySer
V  Q  A  S  L  W  L  Y  L  K  L  L  P  Y  V  L  E  K  G  S

                610                 630                 650
        .         .         .         .         .         .
AGGAGGAAGGTACGGGTCAAGGTGTACTTCCAAGAACAGGGTCACGGAGACAGGTGGAAT
ArgArgLysValArgValLysValTyrPheGlnGluGlnGlyHisGlyAspArgTrpAsn
 R  R  K  V  R  V  K  V  Y  F  Q  E  Q  G  H  G  D  R  W  N

                670                 690                 710
        .         .         .         .         .         .
GTGGTGGAGAAGAAGGTGGACCTGAAACGTAGCGGCTGGCATACCTTTCCCATCACAGAG
ValValGluLysLysValAspLeuLysArgSerGlyTrpHisThrPheProIleThrGlu
V  V  E  K  K  V  D  L  K  R  S  G  W  H  T  F  P  I  T  E

                730                 750                 770
        .         .         .         .         .         .
GCCATCCAGGCCTTGTTTGAGCGAGGGGAGAGACGCCTTAACCTGGATGTGCAGTGTGAC
AlaIleGlnAlaLeuPheGluArgGlyGluArgArgLeuAsnLeuAspValGlnCysAsp
 A  I  Q  A  L  F  E  R  G  E  R  R  L  N  L  D  V  Q  C  D

                790                 810                 830
        .         .         .         .         .         .
AGCTGCCAGGAGCTGGCCGTGGTGCCTGTGTTCGTGGACCCCGGTGAGGAGTCACACAGG
SerCysGlnGluLeuAlaValValProValPheValAspProGlyGluGluSerHisArg
 S  C  Q  E  L  A  V  V  P  V  F  V  D  P  G  E  E  S  H  R
```

```
                850                 870                 890
        .         .         .         .         .         .
CCCTTTGTAGTGGTGCAGGCCCGCCTGGGCGATAGCAGACATCGCATCCGCAAACGGGGC
ProPheValValValGlnAlaArgLeuGlyAspSerArgHisArgIleArgLysArgGly
 P  F  V  V  V  Q  A  R  L  G  D  S  R  H  R  I  R  K  R  G

                910                 930                 950
        .         .         .         .         .         .
CTAGAGTGTGATGGGCGGACCAGCCTCTGTTGCAGGCAACAGTTCTTCATCGACTTTCGG
LeuGluCysAspGlyArgThrSerLeuCysCysArgGlnGlnPhePheIleAspPheArg
 L  E  C  D  G  R  T  S  L  C  C  R  Q  Q  F  F  I  D  F  R

                970                 990                1010
        .         .         .         .         .         .
CTCATCGGCTGGAACGACTGGATCATTGCGCCCACTGGCTACTACGGGAACTACTGTGAG
LeuIleGlyTrpAsnAspTrpIleIleAlaProThrGlyTyrTyrGlyAsnTyrCysGlu
 L  I  G  W  N  D  W  I  I  A  P  T  G  Y  Y  G  N  Y  C  E

               1030                1050                1070
        .         .         .         .         .         .
GGCAGCTGCCCGGCCTATCTGGCCGGGGTCCCTGGCTCAGCTTCCTCCTTCCACACAGCC
GlySerCysProAlaTyrLeuAlaGlyValProGlySerAlaSerSerPheHisThrAla
 G  S  C  P  A  Y  L  A  G  V  P  G  S  A  S  S  F  H  T  A

               1090                1110                1130
        .         .         .         .         .         .
GTGGTGAACCAGTACCGCATGCTGGGCCTGAACCCTGGGCCCGTGAACTCTTGCTGCATC
ValValAsnGlnTyrArgMetLeuGlyLeuAsnProGlyProValAsnSerCysCysIle
 V  V  N  Q  Y  R  M  L  G  L  N  P  G  P  V  N  S  C  C  I

               1150                1070                1090
        .         .         .         .         .         .
CCTACCAAGCTGAGCTCCATGTCCATGCTCTACTTTGATGACGAGTACAACATTGTCAAG
ProThrLysLeuSerSerMetSerMetLeuTyrPheAspAspGluTyrAsnIleValLys
 P  T  K  L  S  S  M  S  M  L  Y  F  D  D  E  Y  N  I  V  K

               1110                1130                1150
        .         .         .         .         .         .
CGGGATGTGCCCAACATGATCGTGGAGGAGTGTGGCTGCGCCTGAcagaggcaacgggggc
ArgAspValProAsnMetIleValGluGluCysGlyCysAlaEnd
 R  D  V  P  N  M  I  V  E  E  C  G  C  A  *

               1170                1190                1210
        .         .         .         .         .         .
ggagcacaggcccatgggtctttgagggagcaggagaggcaggtgggctgagtgtggttg

      1230                1250                1270
        .         .         .         .         .         .
ttccattgggccgtgaagagtccagggtgaggcctgaaataatgttctccgctttgtaga

      1290                1310                1330
        .         .         .         .         .         .
aaaccagtcaggaccagagggagaatccctctgtggcacgagagactcctaacgtcacac

      1350                1370                1390
        .         .         .         .         .         .
atagacacgcatagccagactcacgcagtctgccacccacacagcagcctctggg
```

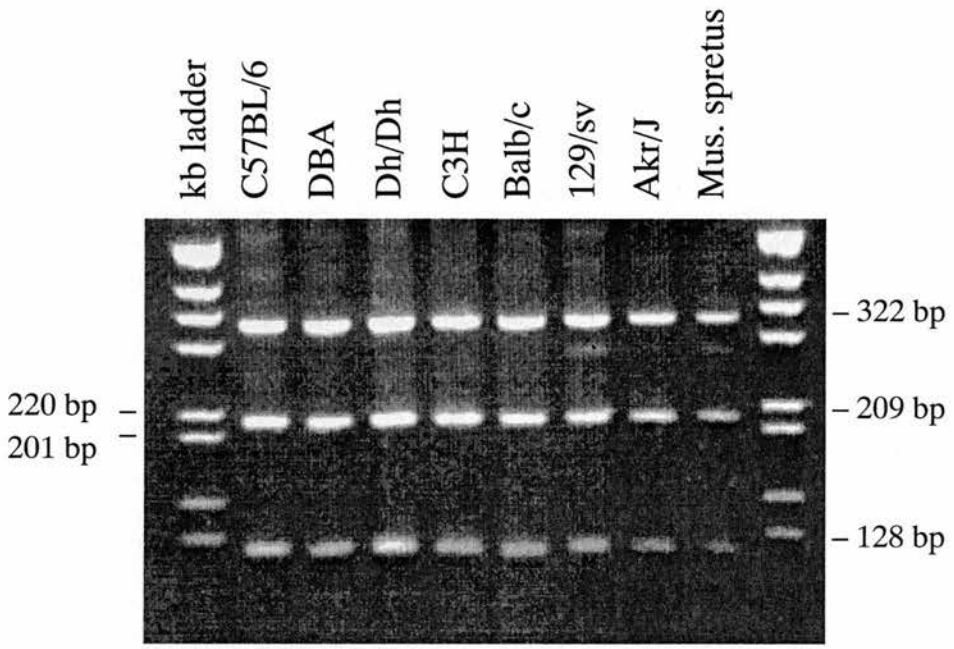



**Figure 4.6 Analysis of sequence variation at nucleotide 1023**

PCR amplification of genomic DNA with primer pair I181/I182 followed by *Hpa*II digestion and resolution on a 5% Nusieve agarose gel. The sequence of Rituos *et al.* (1995) predicts two *Hpa*II sites in the 670 bp fragment, giving digestion products of 322 bp, 220 bp and 128 bp. The sequence variation at nucleotide presented here creates an additional *Hpa*II site which further cleaves the 220bp fragment to 209 bp and 11 bp. Each of the strains analysed produces a fragment of 209 bp, not 220 bp.

## Figure 4.7 Comparison of *Inhβb* peptide sequence from mouse, rat and human

The predicted peptide sequences of the entire *Inhβb* coding region from mouse, rat (Feng *et al.*, 1989, Esch *et al.*, 1987) and human (Mason *et al.*, 1989) are aligned to show maximum homology. A single gap is introduced into the human sequence. Variant amino acids are indicated by *. The site of cleavage resulting in the mature C-terminal *Inhβb* peptide is indicated by ↓.

```
                10        20        30        40        50        60
                .         .         .         .         .
Mouse   MDGLPGRALGAACLLLLVAGWLGPEAWGSPTPPPSPAAPPPPPPPGAPGGSQDTCTSCGG
Rat     MDGLPGRALGAACLLLLAAGWLGPEAWGSPTPPPSPAAPPPPPPPGAPGGSQDTCTSCGG
Human   MDGLPGRALGAACLLLLAAGWLGPEAWGSPTPPPTPAAPPPPPPPGSPGGSQDTCTSCGG
                        *                   *         *

                70        80        90        100       110
                .         .         .         .         .
Mouse   GGGGFRRPEELGRVDGDFLEAVKRHILSRLQLRGRPNITHAVPKAAMVTALRKLHAGKVR
Rat     GGGGFRRPEELGRVDGDFLEAVKRHILSRLQLRGRPNITHAVPKAAMVTALRKLHAGKVR
Human   ....FRRPEELGRVDGDFLEAVKRHILSRLQMRGRPNITHAVPKAAMVTALRKLHAGKVR
                                    *

                120       130       140       150       160
                .         .         .         .         .
Mouse   EDGRVEIPHLDGHASPGADGQERVSEIISFAETDGLASSRVRLYFFVSNEGNQNLFVVQA
Rat     EDGRVEIPHLDGHASPGADGQERVSEIISFAETDGLASSRVRLYFFVSNEGNQNLFVVQA
Human   EDGRVEIPHLDGHASPGADGQERVSEIISFAETDGLASSRVRLYFFISNEGNQNLFVVQA
                                                  *

                170       180       190       200       210
                .         .         .         .         .
Mouse   SLWLYLKLLPYVLEKGSRRKVRVKVYFQEQGHGDRWNVVEKKVDLKRSGWHTFPITEAIQ
Rat     SLWLYLKLLPYVLEKGSRRKVRVKVYFQEQGHGDRWNVVEKKVDLKRSGWHTFPITEAIQ
Human   SLWLYLKLLPYVLEKGSRRKVRVKVYFQEQGHGDRWNMVEKRVDLKRSGWHTFPLTEAIQ
                                            *   *             *

                220       230       240       250       260
                .         .         .         .         .    ↓
Mouse   ALFERGERRLNLDVQCDSCQELAVVPVFVDPGEESHRPFVVVQARLGDSRHRIRKRGLEC
Rat     ALFERGERRLNLDVQCDSCQELAVVPVFVDPGEESHRPFVVVQARLGDSRHRIRKRGLEC
Human   ALFERGERRLNLDVQCDSCQELAVVPVFVDPGEESHRPFVVVQARLGDSRHRIRKRGLEC

                270       280       290       300       310
                .         .         .         .         .
Mouse   DGRTSLCCRQQFFIDFRLIGWNDWIIAPTGYYGNYCEGSCPAYLAGVPGSASSFHTAVVN
Rat     DGRTSLCCRQQFFIDFRLIGWNDWIIAPTGYYGNYCEGSCPAYLAGVPGSASSFHTAVVN
Human   DGRTNLCCRQQFFIDFRLIGWNDWIIAPTGYYGNYCEGSCPAYLAGVPGSASSFHTAVVN
            *

                320       330       340       350       360
                .         .         .         .         .
Mouse   QYRMLGLNPGPVNSCCIPTKLSSMSMLYFDDEYNIVKRDVPNMIVEECGCA
Rat     QYRMRGLNPGPVNSCCIPTKLSSMSMLYFDDEYNIVKRDVPNMIVEECGCA
Human   QYRMRGLNPGTVNSCCIPTKLSTMSMLYFDDEYNIVKRDVPNMIVEECGCA
            *     *           *
```

**Table 4.3 Oligonucleotides used in the analysis of the *Inhβb* coding sequence**

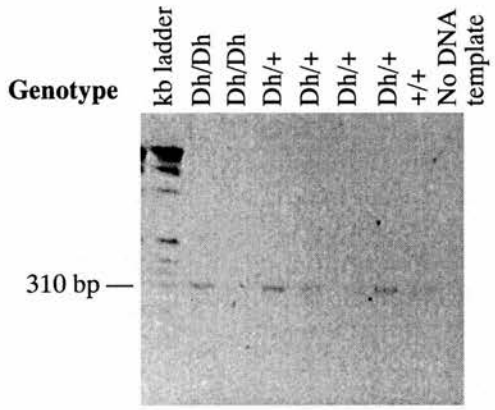
The position of each oligonucleotide is stated relative to the *Inhβb* coding sequence where nucleotide 1 is the first nucleotide of the initiation codon. J116, J95 and J96 are located in non-coding regions and are designated accordingly. RT-PCR, reverse transcription PCR; V-PCR, vectorette PCR; C.S., colony screening; Seq, sequencing.

| Oligonucleotide and nucleotide position | | Sequence 5′ - 3′ (F; forward, R; reverse) | Application |
|---|---|---|---|
| *Exon 1* | | | |
| J116, -67 | F | GGAGCCCGGCCCTGAGCTCG | RT-PCR, Seq |
| I940, 1 | F | ATGGACGGGCTGCCCGGTCG | C.S., Seq |
| I235, 152 | F | CGCAGGACACCTGTACGTCGTG | RT-PCR, V-PCR, Seq |
| I771, 243 | R | CGCCTCCAGGAAGTC(A/G)CC(G/A)TC | RT-PCR, V-PCR, C.S., Seq |
| I941, 439 | F | ATCATCAGCTTTGCAGAGAC | RT-PCR, C.S., Seq |
| J95, 460+30 | R | CGCCTCTGGCTTCCCTGTGG | Seq |
| *Exon 2* | | | |
| I181, 492 | F | CTTCTTCGTCTCTAATGAAGG | C.S., Seq |
| I772, 577 | R | GGACATAGGGGAGCAGTTTCAGG | RT-PCR, V-PCR, Seq |
| I910, 713 | F | CCATCCAGGCCTTGTTTGAG | Seq |
| I908, 898 | R | ACTCTAGGCCCCGTTTGCGG | Seq |
| I909, 951 | F | GGCTCATCGGCTGGAACGAC | Seq |
| I773, 1106 | F | TACCGCATGCGTGGCCTGAACCC | V-PCR, Seq |
| I182, 1162 | R | GAGCATGGACATGGAGCTCAGC | Seq |
| I942, 1237 | R | TCAGGCGCAGCCACACTCCTC | C.S. |
| J96, 1237+162 | R | ATTCTCCCTCTGGTCCTGAC | RT-PCR |

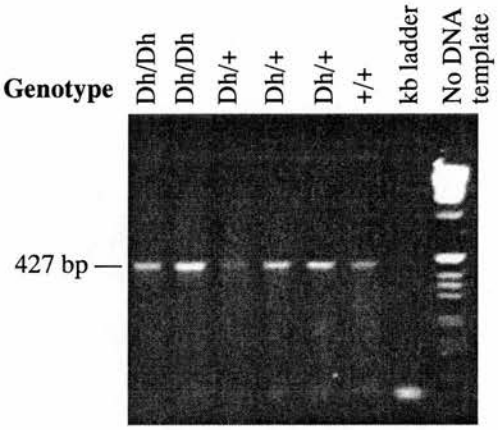
### 4.2.4 Analysis of the *Inhβb* coding sequence in *Dh* mice

To determine whether *Inhβb* is expressed in *Dh* animals and to generate homozygous *Dh Inhβb* coding sequence amenable to sequence analysis, RT-PCR was performed on the *Dh*/+ x *Dh*/+ litter characterised in section 4.2.1. The coding sequence of *Inhβb* was amplified in three fragments using primer pairs J116/I771, I235/I772 and I941/J96 (see table 4.3). This analysis demonstrates the predicted sized fragments to be amplified from wild type and *Dh* heterozygous and homozygous animals (figure 4.8). There is no polymorphism intrinsic to the *Inhβb* sequence to enable the genotype of RT-PCR product to be confirmed. Therefore to control for contamination,*Gli2* primer pair 12a/1b was used to amplify from the same reverse transcription reactions from which the *Inhβb* fragments were amplified. Digestion of the 12a/1b product with *Hpa*II, which is polymorphic between *Dh* and C57BL/6, showed that there was no cross contamination. Although two transcripts of 4.2 kb and 3.4 kb have been demonstrated by Northern blot analysis (Feijen *et al.*, 1994, van den Eijnden-van Raaij *et al.*, 1992), no splice variants were detected. This indicates that *Inhβb* is expressed in *Dh* animals and that there are no gross alterations affecting the coding sequence or splice sites.

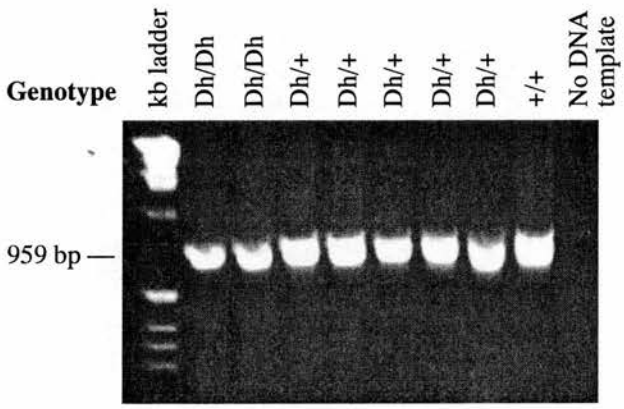
To detect coding sequence mutations RT-PCR product from *Dh* homozygous animals was directly sequenced. Unequivocal sequence was determined from both strands for the majority of the *Inhβb* coding sequence and no sequence variation was detected compared to wild type. However two small regions were refractory to obtaining unambiguous sequence; nucleotides 121 - 138 and 241 - 264.




**J116/I771**




**I235/I772**




**I941/J96**

### 4.3 Discussion

Two genes, *Gli2* and *Inhβb,* have been identified as strong candidates for the *Dh* mutation. The molecular lesion responsible could take one of many forms, however, as discussed in chapter 1 (1.3.3) this must result in a gain of function mutation. The vast majority of gain of function mutations, characterised to date, result from lesion directly affecting coding sequence (see Wilkie, 1994). The mechanisms responsible include large chromosomal rearrangements such as translocation, duplication, large deletion or inversion. Southern analysis of the *Gli2* locus by David Hughes indicated the absence of any such rearrangement in *Dh* animals. This analysis has not been performed for the *Inhβb* locus. Further mutational mechanisms include nonsense or frameshift mutation, missense or small in frame deletion or insertion or splice site mutation. To assay for the presence of such mutation the *Dh* allele of both genes was assessed by a thorough examination of the respective coding regions. The entire coding sequence of the *Dh* allele of each gene was amplified by RT-PCR and the coding sequence comprehensively determined by sequence analysis (two small regions, nucleotides 121-138 and 241-264 remain to be unequivocally sequenced from the *Dh* allele of *Inhβb*). This analysis failed to detect a mutation. The amplification of the predicted sized RT-PCR fragments and wild type sequence across all splice site junctions also indicates the absence of abberantly sized transcripts due to splice site mutation. In addition the ability to amplify *Gli2* and *Inhβb* from day 12.5 total RNA indicates that both genes are expressed in *Dh* animals at this time.

That no mutation was discovered does not rule out the candidacy of either gene for *Dh*. Indeed, despite the most exhaustive mutational analysis, no gene can ever be excluded on the basis that the molecular nature of the mutation could not be determined. It is highly possible that *Dh* is the result of a lesion in a different gene but there remain several possibilities for mutation in *Gli2* and *Inhβb* which have not been assessed. Firstly the untranslated regions (UTRs) of both genes have not been

analysed and secondly a detailed study for aberrant expression has not been undertaken. The UTRs of genes play roles in mRNA stability, regulation of transcript transport and translation. Mutation therein may therefore result in a loss of control of post-transcriptional processing, resulting in incorrect protein levels. For example, dominant mutations in the heterochronic *C. elegans* gene *lin-14* are caused by deletions in the 3′ UTR. This leads to raised protein levels (by an unknown mechanism) causing reexpression of early cell lineages at later developmental stages (Ruvkun *et al.*, 1991). It is very unlikely that a mutation exists in the 3′ UTR of *Gli2* because the proximal extent of the *Dh* region is defined by a polymorphism in the 3′ UTR of *Gli2*, at nucleotide 5410 (stop codon is 4744). Two animals from the *Dh* backcross show informative recombination between this polymorphism and a microsatelite marker in the central region of *Gli2* (see chapter 3, 3.1.3). Taking the assumption that both recombinations occurred at the polymorphism in the 3′ UTR then the formal possibility remains that the mutation could lie in the 629 bp of unanalysed 3′ UTR. The cloned *Gli2* cDNA is approximately 6 kb and includes only 120bp of 5′ UTR (Hughes *et al.*, 1997) while Northern analysis shows the *Gli2* transcript to be 8kb (Hui *et al.*, 1994). Up to 2 kb of 5′ UTR therefore remains uncloned and unanalysed for mutation. Screening of cDNA libraries and/or the use of cDNA end amplification techniques should be employed to address this.

Similarly the UTRs of *Inhβb* warrant mutational analysis. The largest *Inhβb* transcript detected is 4.2kb (Feijen *et al.*, 1994), compared to the coding region of 1.2kb. Subsequent to this work Albano *et al.*, (1993) released sequence of *Inhβb* which extended 3′ of amino acid 260 and included the entire 3′ UTR which is 1.8 kb. The availability of this sequence therefore facilitates the analysis of the *Dh* allele. By extrapolation the 5′ UTR is approximately 1.1kb and again screening of cDNA libraries and/or the use of cDNA end amplification techniques should be employed to clone this region. In addition to sequence analysis of UTRs the use of specific

antibodies should be employed to determine aberrant levels of protein which may be caused by a fault in post-transcriptional processing.

Mutation affecting transcriptional regulatory domains can result in a disturbance of the exquisite controls of mRNA expression that dictate normal cellular distribution, temporal restriction and absolute levels of mRNA. For example promoter mutations in the *C. elegans* sex determining gene *her-1* increase expression levels and result in partial transformation of XX worms into phenotypic males (Trent *et al.*, 1988). Also the lethal yellow mutant at the mouse *agouti* locus ($A^y/A^y$) results from ectopic, ubiquitous expression of the normal protein Miller *et al.*, 1993; Michaud *et al.*, 1993).

Expression analysis of *Gli2* shows that in day 12.5 total RNA there is no gross difference between the levels of expression from the *Dh* or wild type alleles. This is in agreement with semi-quantitative RT-PCR analysis of *Gli2* from day 10.5 *Dh* and wild type total RNA performed by David Hughes. However this analysis does not preclude temporally altered expression at other time points or subtle levels of ectopic expression. Semi-quantitative RT-PCR analysis of *Inhβb* from day 12.5 *Dh* and wild type total RNA was also performed, however this analysis proved inconclusive. To perform a thorough investigation semi-quantitative RT-PCR analysis should be performed using RNA from several stages of development, from day 7.5, when *Gli2* expression is first detected, through to day 14.5. In addition the sensitivity of this technique would be enhanced if the source RNA was derived from only those tissues which display the *Dh* phenotype, for example, from the lateral mesoderm, at day 8.5, or the limb bud at day 10.5-12.5. The most comprehensive way to investigate ectopic or temporally altered mRNA expression is, however, by in situ hybridisation. This enables the determination of expression at the cellular level and also enables quantification of expression. This has been attempted by Laura Lettice for *Gli2* to compare *Dh* and wild type animals but results to date have been inconclusive.

# CHAPTER 5

# ATTEMPTED IDENTIFICATION OF A NOVEL *Gli* FAMILY GENE, A POTENTIAL CANDIDATE FOR *Hemimelic extra toes*

## 5.1.1 Introduction

Paralogous linkage groups on chromosomes 1 and 5 suggest a common ancestry to the genes involved in the mutations *Hemimelic extra toes*, *Hx*, and *Dh*. At the time of this study *Gli2*, a member of the *Gli* family of zinc finger genes, had been identified as a strong candidate gene for *Dh*. This raised the possibility that a novel *Gli* gene was involved in *Hx*. The work presented in this chapter describes degenerate PCR and yeast artificial chromosome based approaches to the identification of such a gene.

## 5.1.2 *Hx* and *Dh* are located in paralogous genetic linkage groups

Evolution of the vertebrate genome is thought to have encompassed numerous events which gave rise to duplicated genetic information. These duplication events appear to be responsible for the generation of homologous genes which constitute the many multi-gene families. Three main mechanisms have been proposed by which genes duplicate (Nadeau, 1991). Tandem duplication produces adjacent gene duplicates due to unequal crossing-over during meiosis (Zimmer *et al.*, 1980). Subsequent rearrangement may increase or decrease the distance between duplicates or distribute duplicates to other chromosomes, however such rearrangements are expected to be rare due to the initial physical proximity of the duplicates and the very low fixation rates of reciprocal translocation (Hedrick and Levin, 1984). Tandomly duplicated genes therefore tend to remain physically linked. A typical example of clustered genes is found on mouse chromosome 8, where eleven esterase loci are located within an 11cM region (Hillyard *et al.*, 1992).

The second mechanism involves the duplication of whole chromosomes via meiotic non-disjunction while the third results in whole genome duplication by the process of tetraploidisation. By comparing the nuclear DNA content of primitive chordates to that of mammals, Ohno (1970) has argued that up to three rounds of tetraploidisation may have occurred in the lineage leading to mammals, the most recent being 250-300 million years ago. The vestiges of large scale duplications are chromosomal regions containing linked genes that have a corresponding region of related linked genes elsewhere in the genome. Within a species, these related regions and the genes therein, are termed paralogous, while homologous genes of common ancestry found in different species are termed orthologous (Lundin 1979).

Based on the criteria of sequence homology and in some cases conserved function Lundin, (1989, 1993), Nadeau and Kosowsky, (1991) and Nadeau, (1991) have identified paralogous genes within the mouse and human genomes. In addition they have investigated the chromosomal distribution of these genes and demonstrated numerous paralogous linkage groups which testify to duplication events. A typical example is suggested between mouse chromosomes 6 and 7. The *ras*-like oncogene *Kras2* is closely linked to lactate dehydrogenase *LDh1* on mouse chromosome 6 whereas the related genes *Hras1* and *LDh2* are tightly linked on chromosome 7. Corroboration and extension of this region has recently been shown by the localisation of the neurotrophic factor gene *Ntf3* and the potassium channel genes *Kcna1*, *Kcna5*, *Kcna6*, *Kcnalrs1* to chromosome 6 and the paralogous loci *Ntf5* and *Kcnalrs4* to chromosome 7 (Lock *et al.*, 1994).

There are several factors which obscure the evidence for chromosomal or genome duplication events and which complicate the elucidation of regions of paralogy. Obviously physical loss of a gene removes evidence for duplication. Similarly if a gene becomes non-functional soon after duplication it is unlikely that the paralogous gene could be identified due to the high rate of sequence divergence which occurs without functional constraint (Li, 1982). Also chromosomal rearrangements may disrupt duplicated regions which will decrease the size of

paralogous regions but increase their number. However the rapid progress being made in sequencing genes and the mapping of loci to mammalian genomes, is enabling ever more refined and meaningful comparisons to be made between chromosomal regions within and across species.

The genetic location of *Dh* and *Hx*, in relation to flanking loci, suggests that the two loci may reside in paralogous linkage groups. As illustrated in figure 5.1, *Dh* maps 0.7cM from *En1* on chromosome 1 while *Hx* maps 1.1cM from *En2* on chromosome 5 (Martin *et al.*, 1990, Higgins *et al.*, 1992). Also assigned to these regions are the gastrulation brain homeobox genes, *Gbx1* and *Gbx2*. Frohman *et al.* (1993) observed no recombination between *Gbx1* and *En2* and between *Gbx2* and *En1* following analysis of 28 and 30 meiosies respectively of an interspecific backcross. Subsequently analysis of a 636 animal backcross by Montgomery *et al.* (1994) has positioned *Gbx1* 1.9cM proximal to *En2*. Most recently the paralogous seretonin receptor genes *Htr5a* and *Htr5b* have been mapped, also by interspecific backcross analysis in the vicinity of *Hx* and *Dh* respectively (Danielson *et al.*, 1994; Matthes *et al.*, 1992). That three unrelated, tightly linked loci on chromosome 1 have paralogous loci tightly linked on chromosome 5 strongly suggests that these regions originated following an ancient chromosomal duplication event.

Figure 5.1 also shows the map positions of the ecotropic proviral loci *Emv1* and *Emv17*, linked to *Hx* and *Dh* respectively. The proviral sequences themselves cannot be part of the paralogous groups because they are present in some strains of laboratory mice but not others (Jenkins *et al.*, 1982). However there is some evidence that preferential sites for virus integration exist (Shih *et al.*, 1988). It is therefore possible that such a site was present in the ancestral chromosome. It may therefore be valid to consider the *Emv1* and *Emv17* loci as part of the paralogous linkage groups.

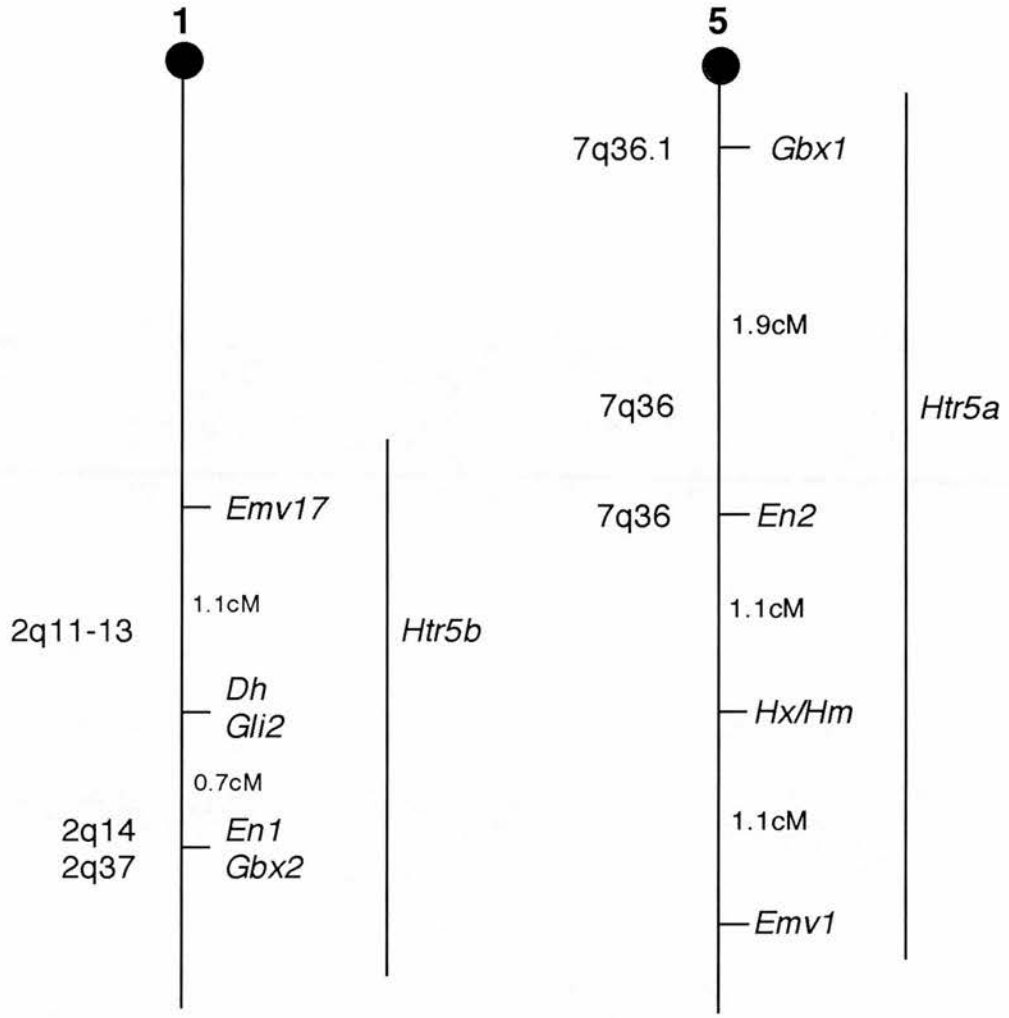



**Figure 5.1 Consensus maps of paralogous loci linked to *Dh* and *Hx* on chromosomes 1 and 5**

Genetic distances between loci are shown in cM. The map positions of *Htr5a* and *Htr5b* relative to the loci shown have not been determined. Their locations within the linkage groups are therefore not known, as indicated by the dashed lines. Map positions for human orthologues are shown to the left of each locus.

Supporting evidence for the paralogy of these regions is provided by the comparative map of the human genome and indicates a duplication event prior to the divergence of humans and mice. With the exception of *GBX2*, the orthologous loci of these genes are also found to be tightly linked in the human genome, on chromosomes 2q and 7q (figure 5.1). The location of *GBX2* on distal chromosome 2 is likely to

reflect intra-chromosomal rearrangement following the original duplication event (Matsui *et al.*, 1993).

It is possible that the location of *Dh* and *Hx* relative to these regions of paralogy is coincidental, the two genes not being related. However since *Dh* and *Hx* are located at similar genetic distances from *En1* and *En2* respectively and *Dh* and *Hx* exhibit a similar limb phenotype, as discussed in chapter 1 (1.3) it is highly possible that the two alleles are paralogous genes within these linkage groups. Thus an ancient duplication event would have encompassed not only the ancestral *En*, *Gbx* and *Htr5* genes but also a gene required for normal limb morphogenesis. Consequently *Dh* and *Hx* would be mutations in paralogous genes.

### 5.1.3 A novel *Gli* gene is a potential candidate for *Hx*

If *Dh* and *Hx* are alleles of paralogous genes then identification of a candidate gene for either mutation will mean a homologous gene is a candidate for the other. At the time of this study *Gli2* had been identified as a strong candidate for *Dh*, see chapters 3 and 4. This therefore implicated a *Gli* gene as candidate for *Hx*. The mammalian *Gli* family currently extends to three members, *Gli*, *Gli2* and *Gli3*. *Gli* has been mapped to chromosome 10 (Justice *et al.*, 1990) and *Gli3* to chromosome 13 (Hui and Joyner, 1993). *Hx* would therefore represent an allele of a novel *Gli* gene.

To investigate the existence of such a novel *Gli* gene two methodologies were employed. The *Gli* family is characterised by five tandem zinc finger motifs of the $C_2H_2$ subclass which are related to but distinct from the Krüppel family of zinc finger genes. This domain is highly conserved between the known mammalian *Gli* genes. At the amino acid level *Gli2* and *Gli3* are 92% identical to each other and 84 % identical to *Gli* (Ruppert *et al.*, 1988, Hughes *et al.*, 1997). This domain can therefore be exploited by the technology of degenerate PCR to amplify related genes. The chromosomal location of novel genes isolated in this way can then be determined and related to the map position of known mouse mutants. Should any novel *Gli* gene map

in the vicinity of *Hx* and the tightly linked mutant *Hammertoe*, *Hm*, (see chapter 1, 1.3.4) then it can be regarded as a candidate for these mutants.

The second methodology employed was to use YAC technology to clone the genomic region which may contain a *Gli2* paralogue. At the time of this study previous analysis of the *Dh* region on chromosome 1 had suggested a physical linkage between *En1* and *Gli2*. PFGE analysis of *BssH*II and *Mlu*I digested genomic DNA had indicated *En1* and *Gli2* probes hybridising to the same 600kb fragments (data not shown). This therefore placed the two genes a maximum of 600kb apart. Should a *Gli2* paralogue be linked to *En2* then the physical distance between the two is likely to be similar to that between *En1* and *Gli2*. The aim of this work was therefore to clone at least 600kb of genomic DNA which extends distally from *En2*, into the *Hx/Hm* region. This can be achieved through the isolation of YAC clones which typically contain large genomic fragments of 500kb-1Mb. Primarily such YACs can then be analysed for the presence of a novel *Gli* gene, however they also provide the resource to identify additional genes from the *Hx/Hm* region. These genes could then be assessed as possible candidates for *Hx/Hm* and exploited to determine the existence of paralogues located either in the *Dh* region of chromosome 1 or elsewhere in the genome. Such data would refine the extent of these regions of paralogy and in addition may identify novel candidate genes for *Dh*.

## 5.2 Results

### 5.2.1 Screening for novel *Gli* genes by degenerate PCR technology

#### Design of degenerate primers specific for *Gli* genes

Many applications of the PCR depend on the availability of DNA sequence from the locus of interest. From this information unique oligonucleotide primers can be designed to enable amplification of specific DNA fragments. There are however occasions when PCR amplification of a particular sequence is desired but only protein sequence is known. The degeneracy of the genetic code precludes the simple translation of amino acid sequence to nucleotide sequence but this limitation can be overcome by the use of 'degenerate' PCR primers. Such primers are a mixture of oligonucleotides which correspond to either a subset, or all of the possible nucleotide sequences which could code for a particular stretch of peptide. Under appropriate reaction conditions a proportion of these oligonucleotides will be sufficiently homologous to prime amplification from the desired sequence(s), while non-specific amplification remains minimal (for practical considerations of the technique see McPherson *et al.*, 1991).

The technique has found widespread application, particularly in gene cloning strategies. For example peptide sequence, generated from an isolated protein, can be used to design PCR primers which are sufficiently degenerate to enable amplification from the encoding gene. The fragment generated can then be used to isolate the entire gene by screening cDNA or genomic libraries (for example, Lee *et al.*, 1988, have cloned the urate oxidase gene using this rational). This methodology has largely superseded the technically difficult procedure of 'guessmer screening'; library screening by hybridisation using a pool of oligonucleotides degenerate to a particular peptide sequence (Sambrook *et al.*, 1989).

The technique has also been extensively used to clone novel members of multigene families. Degenerate primers can be designed to the consensus peptide sequence of the domain, or domains which define such families. Any genes which

contain this peptide sequence and by assumption the corresponding conserved domain are reasoned to amplify during the PCR.

As gene families extend across phyla, homologous genes from diverse species can be identified, for example Chang *et al.* (1994) identified genes of the *hedgehog* family in the flour beetle, leech, sea urchin, mouse and man. Similarly novel genes belonging to a gene family previously defined within a species can be discovered, for example Hötten *et al.* (1995) have cloned a novel human activin/inhibin gene, *inhibinβc*, by employing primers designed from the sequence of the human *inhibinβa* and *inhibinβb* genes. Because this technique is a powerful and rapid approach to the identification of novel multigene family members I have applied it with the intent of identifying novel members of the *Gli* gene family.

There are several conditions that need to be considered in the design of degenerate primers for PCR. Firstly it is desirable to keep the overall degeneracy of a set of primers to a minimum. A high degree of redundancy will increase the number of spurious target sequences to which a primer could anneal and thus reduce specificity. In this respect it is important to examine the amino acid content of a region to be amplified because of the variable degeneracy of the genetic code, (Table 5.1). Where possible selection of sequence should avoid amino acids of high redundancy, such as lysine, serine and arginine.

**Table 5.1 Codon Degeneracy**

| Amino acid | Number of codons |
|---|---|
| M, W | 1 |
| C, D, E, F, H, K, N, Q, Y | 2 |
| I | 3 |
| A, G, P, T, V | 4 |
| L, R, S | 6 |

Primer degeneracy can be greatly reduced by the use of the universal nucleotide inosine in positions of high, 3 or 4 fold, redundancy (Knoth *et al.*, 1988). McPherson *et al.*, (1991) have successfully used inosine to reduce primer redundancy from 73,728 fold to just 72 fold. This allows greater flexibility in choosing amino acid sequence because the highly degenerate amino acids such as serine, leucine and arginine can successfully be included in primer design. A further consideration in terms of reducing primer degeneracy is that of preferential codon usage. It is observed that many species greatly favor the use of only a subset of the possible codons which code for a particular amino acid. (Nakamura *et al.*, 1996). In particular cases certain codons may therefore be ignored.

The 3′ end of any PCR primer is crucial to the performance of the reaction, it being essential that at least the 3′ most nucleotide is complementary to the template. Sommer and Tautz, (1989) have shown that for an efficient reaction giving a high product yield the last 3 nucleotides need to be complementary and should not contain inosine. This may be ensured by ending primers with methionine or tryptophan, which are coded for by only one codon. Primer length will often be dependent on the amino acid sequence available, but where possible should be sufficient to enable the use of a stringent annealing temperature. This will reduce non-specific amplification.

The complete or partial protein sequences of five mammalian *Gli* genes were available at the time of this study; human *Gli*, *Gli2*, *Gli3* and mouse *Gli* (Kinzler *et al.*, 1988, Ruppert *et al.*, 1988) and *Gli2* (Hughes *et al.*, 1997). The defining characteristic of *Gli* genes are 5 zinc fingers which have the consensus sequence [Y/F]XCX$_3$GCX$_3$[F/Y]X$_5$LX$_2$*HX*$_{3-4}$H[T/S]GEKP (Ruppert *et al.*, 1988). This is the most highly conserved region between the genes, *Gli2* and *Gli3* are 92% identical to each other and 84 % identical to *Gli*. This was therefore the region chosen from which to design degenerate primers. The sequences of the zinc finger region were aligned, as shown in figure 5.2, to determine the optimal sequences from which PCR primers should be designed.

One characteristic region of the zinc fingers is the nine amino acid H-C link, which joins the histidine of one finger to the cysteine of the next. This sequence, H[T/S]GEKP[Y/F]XC, is a common characteristic of mammalian Krüppel-like zinc finger genes which are estimated to number approximately 300 (Bellefroid *et al.*, 1989). To avoid the amplification of these genes this sequence was avoided. Four primers were designed according to the criteria discussed above and are shown in Table 5.2. The primers are all fully degenerate for the amino acid sequence, no allowance being made for preferential codon usage. E195 and E196 were designed for amplification from cDNA while E187 and E197 were designed for amplification across a single exon to enable genomic DNA to be used as template. The position of the primers relative to the zinc fingers is shown in figure 5.2.

**Figure 5.2 *Gli* family zinc finger structure.**

Aligned amino acid sequences of the zinc finger region of the 5 reported mammalian *Gli* genes. Mouse *Gli* genes are designated *MGli* and *MGli2* Amino acids are invariant except where denoted by an asterisk. Arrowheads indicate the position of intron/exon junctions. The positions of the degenerate primers, E187, E195, E196 and E197 are indicated, the amino acid sequence shown in bold.

**Table 5.2 *Gli* specific degenerate PCR primers**

The amino acid sequence from which each primer is designed is shown above the deduced primer sequence. In the oligonucleotide sequence, I represents inosine.

| | |
|---|---|
| **F187**<br>Finger 1, (forward) 5′-3′ | H I N N/S D/E H I H G<br>CA(T/C) ATI AA(T/C) A(G/A)I GAI CA(T/C) ATI CA(T/C) GG |
| **E195**<br>Finger 2, (forward) 5′-3′ | A Q Y M L V V H M<br>GCI CA(A/G) TA(T/C) ATG (C/T)TI GTI GTI CA(T/C) ATG |
| **E197**<br>Finger 2, (reverse) 5′-3′ | H R R M H V V L M<br>TG I(T/C)G I(T/C)G CAT (A/G)TG IAC IAC IA(G/A) CAT |
| **E196**<br>Finger 5, (reverse) 5′-3′ | P G H V T K V H K<br>GG ICC (A/G)TG IAC IGT (T/C)TT IAC (A/G)TG (T/C)TT |

### 5.2.2 Restriction enzyme 'fingerprint' analysis of *Gli* genes amplified by degenerate PCR from cDNA template

Due to conservation of the zinc finger region it would be expected that amplification from novel *Gli* genes would produce fragments of the same or very similar size to those of the other *Gli* genes. Fragment size cannot therefore be used to differentiate between the different genes. The nucleotide sequence would however differ, resulting in restriction site variation. As described by Boehm, (1993) this can be exploited in the analysis of the different gene family members by a restriction enzyme fingerprinting procedure.

The PCR is performed to generate product which is radiolabelled at one end only. This is achieved by radiolabeling only one of the primers. The product can then be digested with a panel of frequent cutting restriction enzymes, separated by gel electrophoresis and visualised by autoradiography. This analysis on a cloned gene would reveal one band per digest, the size of the band depending on the position of the restriction site nearest the labeled end of the fragment. However when performed

on a multigene family amplified from genomic or cDNA a complex pattern of bands would be expected due to differences in the position of restriction sites in different genes. This procedure can produce an overall impression of the complexity and approximate gene number within a family. Further, visualised bands can be isolated and used to clone the corresponding genes.

Amplification was carried out using primers E195 and E196 to generate fragments which span zinc fingers 2-5. Random primed cDNA (prepared by David Hughes) from 12.5 day total RNA was used as template. 50μl reactions were set up consisting of the following components: 2mM each primer, 1.5mM $MgCl_2$, 1x PCR buffer (10mM Tris-HCl pH8.3, 50mM KCl), 2μM each dNTP, 2U Taq polymerase, 50ng cDNA. Reaction conditions: 94°C for 3min x 1 cycle, 94°C for 45sec, 45°C for 45sec, 72°C for 1min x 35 cycles, 72°C for 5 min x 1 cycle. These conditions vary from the standard in two ways. Firstly a high concentration of primer is used, to allow for the subset which will not significantly perform in the reaction. Secondly a low annealing temperature of 45°C is used to encourage amplification from divergent sequences. Although it is often necessary to empirically optimise reaction conditions when using degenerate primers, these initial parameters efficiently amplified only fragments of the expected size, 318bp, as shown in figure 5.3.

To perform the fingerprinting procedure the PCR was repeated as above, except that approximately one tenth of the primer E195 was radiolabelled at the 5′ end using T4 polynucleotide kinase and [$\gamma$-$^{32}$P]ATP. The integrity of the amplified DNA was initially checked by analysis of an aliquot by agarose gel electrophoresis. The DNA was then purified by ethanol precipitation followed by resuspension in 50μl T.E. buffer. 5μl was then digested with each of the following enzymes; *Alu*I, *Hae*III, *Msp*I, *Rsa*I, *Sau*3AI and *Taq*I. Samples were then separated on a 6% non-denaturing polyacrylamide gel and visualised by autoradiography. The result is shown in figure 5.4.

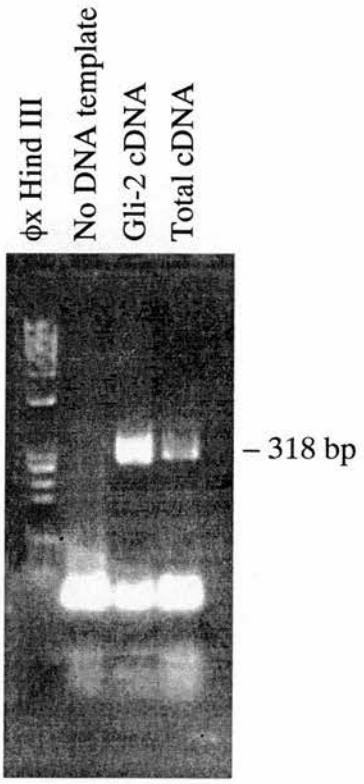



**Figure 5.3 Degenerate PCR amplification of *Gli* genes from cDNA template**

Degenerate PCR amplification of *Gli* genes using primers E195/E196 on cDNA template. The expected size product, 318bp, is amplified from cloned *Gli2* cDNA. The same sized product is amplified from total cDNA (synthesised from E12.5 total RNA). No other predominant products are produced. The lower band in both lanes most likely represent primer-dimer product as they are also generated in the absence of DNA template.

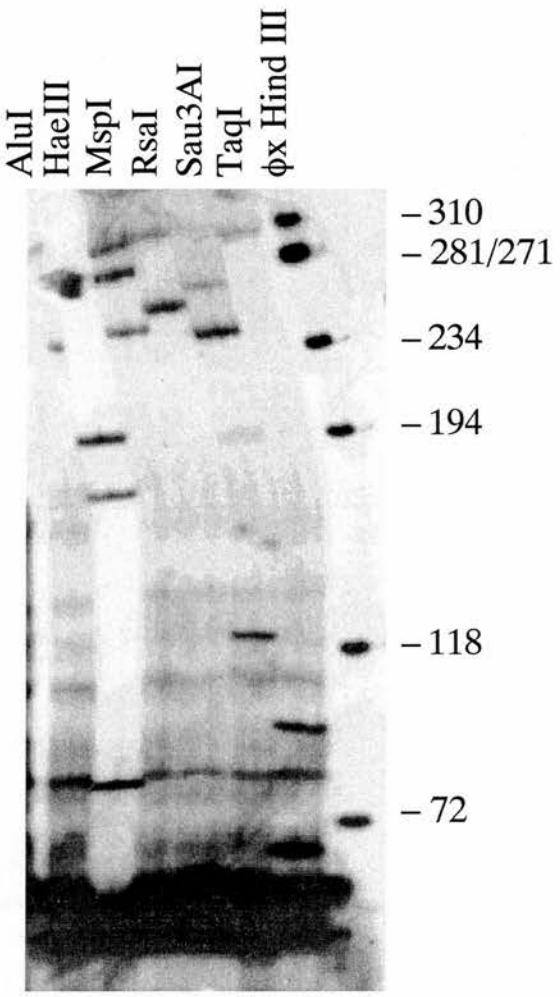



**Figure 5.4 Fingerprint analysis of degenerately amplified *Gli* genes**

6% non-denaturing polyacrylamide gel of restriction digested, asymmetrically labelled amplification products (primers E195/E196) derived from E12.5 total cDNA.

Between two to four bands can be seen in each lane (see also table 5.3). Four bands are seen for the *Sau*3A1 digest, two of which are at a relative low abundancy. These bands could represent four separate *Gli* genes which show different levels of expression. However although attention was paid to achieving complete digestion, the two fainter bands could represent partially digested fragments. The presence of two to three bands produced by the other digestions suggests a gene family of at least three members. Although four or more discrete bands are not produced by these digestions novel *Gli* genes may still be represented if restriction sites are conserved between family members.

To analyse further the genes which had been amplified the DNA from each band was recovered, PCR amplified and cloned to enable individual fragments to be sequenced. Using the autoradiograph as a guide the areas of the gel containing each band were excised. The gel was re-exposed to check that the gel slices recovered did indeed contain the DNA band. The 'crush and soak' method (Sambrook *et al.*, 1989) was used to isolate the DNA which was then treated with DNA polymerase I to produce blunt ended termini. The linker 962L/963LB was then ligated onto each sample to enable PCR amplification using the primer 962L. The termini of amplified DNA were then made blunt using DNA polymerase I and cloned into the *Eco*RV site of pBluescript.

Individual white colonies were subsequently picked and analysed for the correct size of insert by PCR using the pBluescript vector primers T3 and T7. Amplified inserts of the expected size were then sequenced from both ends using T3 and T7 as primers. The number of clones analysed from each excised band is listed in table 5.3. The sequence generated for each fragment was then compared to the available protein and nucleotide sequences of the three mouse *Gli* genes using the Map and Bestfit programs available through the GCG Wisconsin computing package. The identity of all fragments sequenced are shown below in table 5.3.

**Table 5.3 Sequence analysis and identity of fingerprint restriction fragments.**

| Enzyme | Size of bands generated, bp | Number of clones sequenced | Identity of sequenced clones | |
|---|---|---|---|---|
| | | | Identity | Number |
| Alu1 | 280 | 0 | - | - |
| | 249 | 2 | *Gli* | 2 |
| HaeIII | 192 | 2 | *Gli3* | 2 |
| | 180 | 2 | *Gli* | 2 |
| | 83 | 3 | *Gli2* | 3 |
| MspI | 318 | 1 | *Gli2* | 1 |
| | 288 | 2 | *Gli3* | 2 |
| | 254 | 2 | *Gli* | 2 |
| RsaI | 318 | 1 | *Gli3* | 1 |
| | 260 | 4 | *Gli* | 3 |
| | | | *Gli2* | 1 |
| Sau3AI | 275 | 3 | *Gli2* | 3 |
| | 245 | 3 | *Gli* | 1 |
| | | | *Gli2* | 2 |
| | 194 | 2 | *Gli3* | 2 |
| | 120 | 2 | *Gli3* | 2 |
| TaqI | 90 | 3 | *Gli3* | 3 |
| | 67 | 0 | - | - |

As shown above all of the fragments sequenced corresponded to either *Gli*, *Gli2* or *Gli3*. No novel *Gli* gene was detected. Sequence from two of the fragments, *Alu*I 280bp and *Taq*I 67bp, was not determined as clones were not obtained. Two bands consist of more than one gene; *Rsa*I 260bp represents *Gli* and *Gli2* and *Sau*3AI 245bp also represents *Gli* and *Gli2*. Fragments sequenced from the two least abundant bands generated by the *Sau*3A1 digest are indeed shown to be partial digestion products; the 275bp band represents a partially digested *Gli2* fragment and the 194bp band a partially digested *Gli3* fragment because fragments of 269bp and 120 bp are observed for *Gli2* and *Gli3* respectively. 32 independent fragments were sequenced which represent a random sample from the degenerate PCR amplification. Each of the three *Gli* genes were identified approximately the same number of times;

*Gli* 10, *Gli2* 10 and *Gli3* 12 times. This correlates well with the band intensities for each gene which are very similar. The primers E195/E196, under the PCR conditions used are *Gli* specific and are unbiased in their amplification of all known members of the *Gli* family. This would suggest that any novel *Gli* gene should also have been amplified. This analysis may however have missed such a gene for two reasons. The gene may not be expressed at 12.5 days resulting in very little or no transcript in the RNA used to prepare the cDNA template. Secondly divergence in the peptide sequence on which the degenerate primers were based, particularly at the 3′ end of either primer may have caused amplification to fail. To circumvent these two caveats a second set of degenerate primers was designed which would amplify from genomic DNA.

### 5.2.3 Sequence analysis of *Gli* genes amplified by degenerate PCR from genomic DNA

The intron/exon structure of the *Gli* zinc finger region has been determined by Ruppert *et al.*, (1988), see figure 5.2, however the intron sizes have not been determined. PCR amplification across an intron of indeterminate size may well fail if the intron is too large. To use genomic DNA as PCR template it was therefore necessary to design the primers F187 and E197 which would amplify from a single exon.

The PCR conditions used for F187/E197 were the same as for E195/E196, except that 50ng of mouse genomic DNA (C57BL/6) was used in place of cDNA and neither primer was radiolabelled. A band of the expected size, 123bp, was successfully amplified as shown in figure 5.5. The fingerprint analysis was not performed on these amplified fragments because the previous analysis indicates that the mammalian *GLI* gene family is limited to a small number of members. A more rapid approach for the analysis of amplified fragments was to directly clone the PCR product and sequence a random selection of clones. The 123bp band was thus

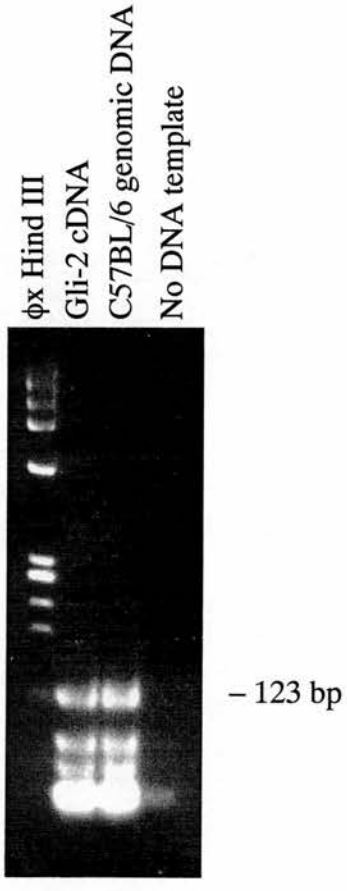



**Figure 5.5 Degenerate PCR amplification of *Gli* genes from genomic DNA template**

Degenerate PCR amplification of *Gli* genes using primers F187/E197 on genomic (C57BL/6) DNA template. The expected size product, 123bp, is amplified from cloned *Gli2* cDNA. The same sized product is amplified from genomic DNA. The lower bands in both lanes most likely represent primer-dimer products as they are also generated in the absence of DNA template.

isolated following agarose gel electrophoresis, cloned into pBluescript and inserts sequenced following the procedure described above for the isolated fingerprint bands.

Sequence from sixteen individual clones was generated and analysed as described above. Ten of these clones were identical to *Gli* while six were identical to *Gli2*. Neither *Gli3* nor a novel *Gli* gene were detected. Although *Gli* specific, the reaction, under the conditions used, was clearly not amplifying the known *Gli* genes at an equal rate. It is important to eliminate this selective bias to increase the chance of amplifying all putative members of the gene family.

The experiment was therefore repeated using the same reaction components as above and a PCR cycling programme with an initial step of five cycles at reduced stringency; 94°C for 3min x 1 cycle; 94°C for 45sec, 37°C for 45sec, 72°C for 1min x 5 cycles; 94°C for 45sec, 50°C for 45sec, 72°C for 1min x 30 cycles; 72°C for 5 min x 1 cycle. The step of five cycles with an annealing stage of 37°C should permit amplification from a greater range of target sequences. This may reduce specificity for the *Gli* family but will aid in the amplification of any potentially divergent family members.

The reaction, under these conditions again produced a single band of 123bp, similar to that shown in figure 5.5. This band was isolated and analysed as above. Thirty two clones were sequenced thirty one of which were identical to *Gli* gene sequences; *Gli*, 10; *Gli2*, 10; *Gli3*, 11. No novel *Gli* gene was detected. A single clone was found to be artifactual.

### 5.2.4 Long range characterisation of the genomic region surrounding *En2*; Isolation of *En2* containing YAC clones

Three YAC libraries were screened to isolate clones containing the marker *En2*. The St. Marys' Hospital library (Chartier *et al.*, 1992) and the Whitehead/MIT library (Kusumi *et al.*, 1993) were screened by PCR using the *En2* specific primers D352/D353. The ICRF library (Larin *et al.*, 1991) is available as gridded filters and was screened by hybridisation using an *En2* specific probe. Library details and screening protocols are presented in the materials and methods chapter, 2.6.2-2.6.3.

Screening of the primary pools from the St. Marys' Hospital and Whitehead/MIT libraries identified one and two positive secondary pools respectively. Screening of the positive St. Marys' Hospital secondary pool identified one positive clone, D9P8. A positive clone, 98D8, was successfully isolated from one of the Whitehead/MIT secondary pools, but the other pool failed to yield appropriate coordinates. Examples of these screening procedures are illustrated in chapter 3, figures 3.6 and 3.7.

Due to the low return from this screen (2 clones from approximately eight genome equivalents) the screen was repeated using a different set of *En2* specific PCR primers, D447/D448. However this assay also identified D9P8 and 98D8 but no additional clones. They did not identify the Whitehead/MIT secondary pool which failed to yield a positive clone, suggesting a false positive result.

The ICRF library was screened with the *En2* specific probe, pλ8SR (Davidson *et al.*, 1988) producing seven potential positive signals. A portion of one of the library filters showing three signals is shown in figure 5.6. Subsequent PCR analysis of the clones corresponding to the seven sets of coordinates demonstrated only one clone, ICRFy903CO22, (subsequently refered to as CO22) to contain *En2*. However at a later date it transpired that an error at the library database had been made in the translation of filter coordinates to clone identity. Three of the seven

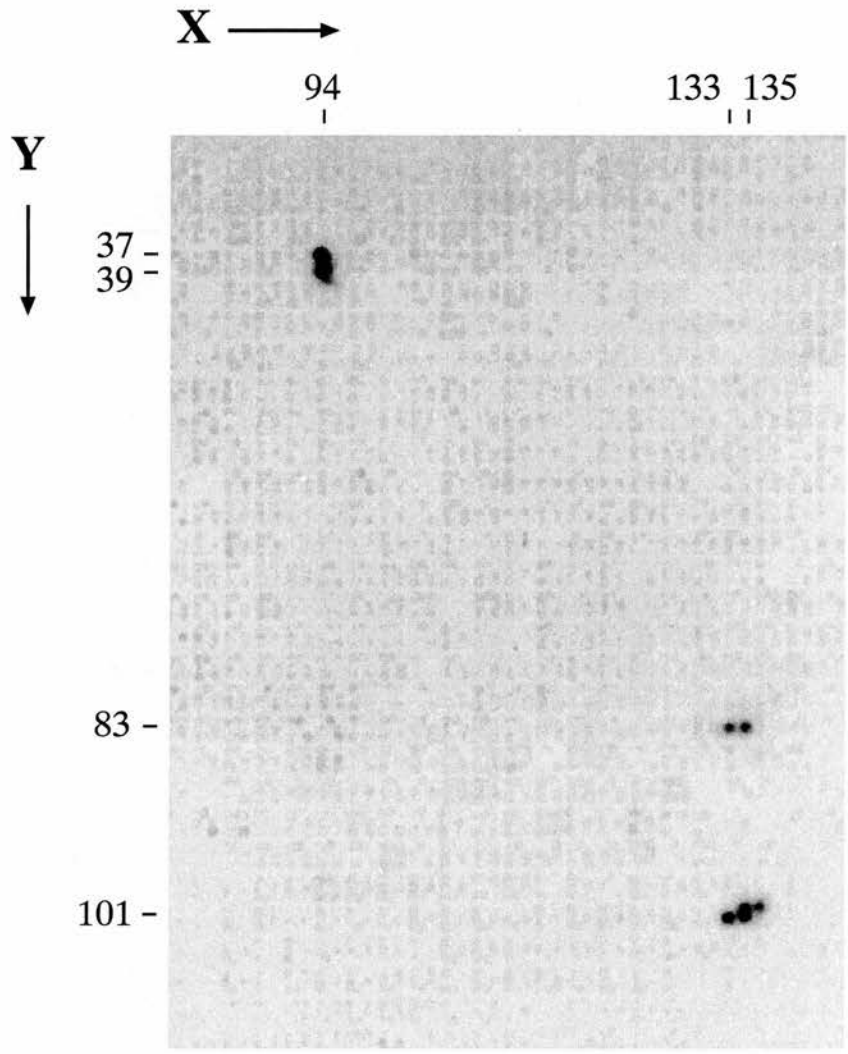



**Figure 5.6 *En2* screen of ICRF YAC library**

Portion of one ICRF high density gridded mouse YAC filter screened with the *En2* specific probe pλ8SR. The grid coordinates, which identify individual clones, are labelled on the X and Y axies for the three positive signals shown. Clones have been spotted in blocks of 3 × 3 clones. This pattern can be seen due to the background identification of all clones by $S^{35}$ labelled AB1380 DNA. To aid identification of true positive clones each clone is represented twice within a block, in a specific orientation. The signal at X = 133,135 Y = 83 identified YAC ICRFy903CO22.

clones received did not therefore correspond to the hybridisation signals produced. The correct clones have not been tested for authenticity.

### 5.2.5 Physical characterisation of YAC clones

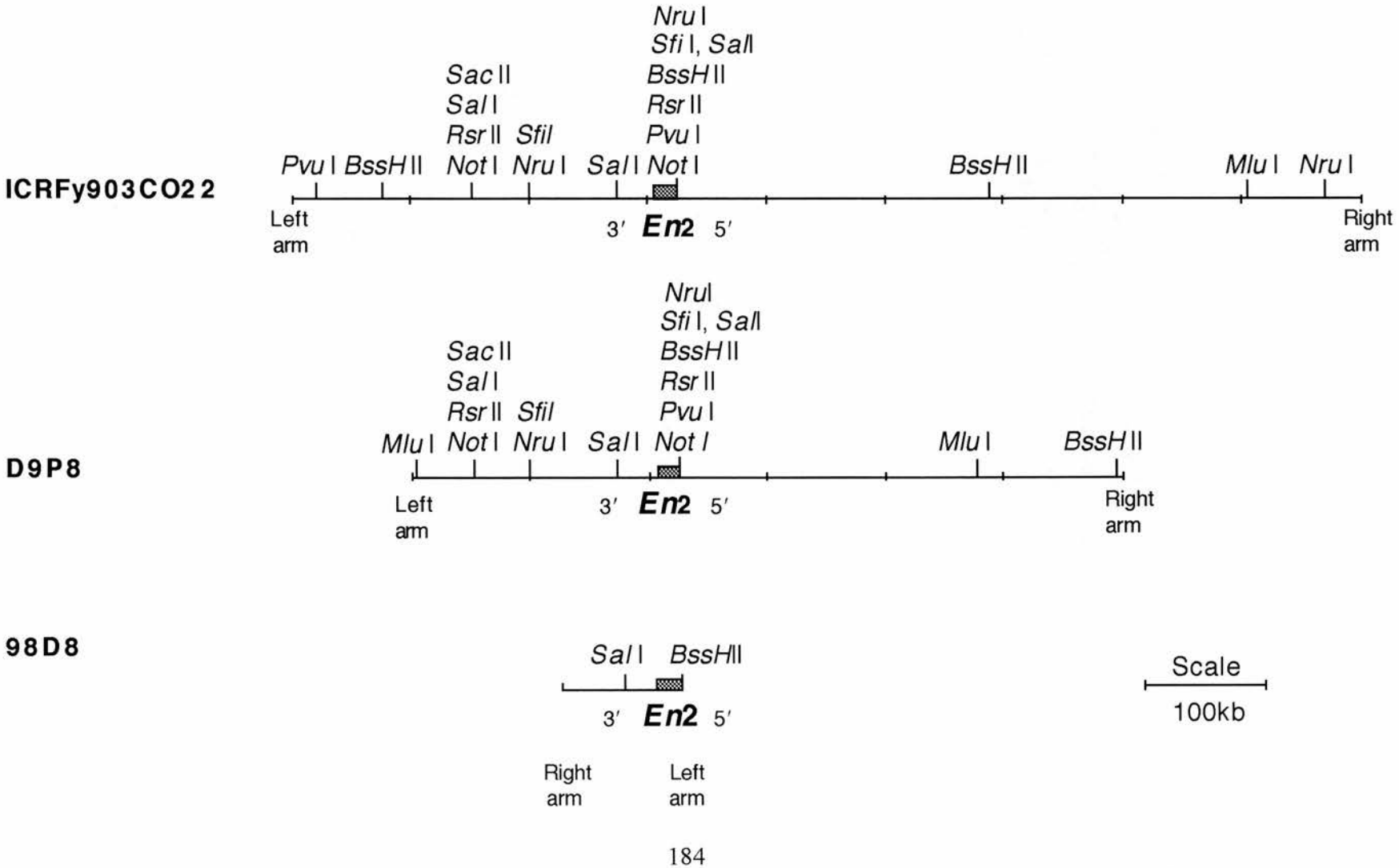
Initial analysis to determine the size of each YAC (see chapter 3, 3.2.7) demonstrated all three clones to contain a single, stable YAC, the sizes of which are CO22, 900kb; D9P8, 600kb and 98D8, 90kb. Restriction maps were generated for each YAC by PFGE of partially restricted YAC DNA according to the procedure described in chapter 3 (3.2.7).

The data produced from each YAC was assimilated into restriction maps. The three maps are shown aligned relative to *En2* in figure 5.7. The position of *En2* within each YAC was determined by assessing the fragments to which the *En2* probe pλ8SR (Davidson *et al.*, 1988) hybridised relative to the restriction map deduced using the vector arm probes. In both CO22 and D9P8 pλ8SR hybridised to the following fragments; *Not*I, 175kb; *Rsr*II, 175kb; *Sac*II, 175kb; *Sal*I, 175kb and 50kb *Nru*I, 125kb; *Sfi*I, 125kb; *BssH*II, 10kb. For CO22 a 300kb *Pvu*I fragment was seen compared to a 225kb fragment seen in D9P8. This pattern is consistent with *En2* being located adjacent to the cluster of sites 325kb and 225kb from the left arms of CO22 and D9P8 respectively. These sites represent the CpG island which has previously been shown to be associated with *En2* (Logan *et al.*, 1992). pλ8SR is a cDNA fragment located downstream of the CpG island which will therefore detect fragments to the 3′ side of this island. Consistent with this the *En2* hybridising fragments all map to one side of the island, towards the left ends of CO22 and D9P8. This also determines the transcriptional orientation of *En2*, as indicated in figure 5.7, although a 5′ *En2* probe was not available to confirm this.

A second cluster of sites including *Not*I and *Sac*II is located 175kb 3′ of *En2*. 89% of *Not*I sites and 74% of *Sac*II sites are estimated to reside within CpG islands (Lindsay and Bird, 1987). This is thus likely to represent a CpG island associated

**Figure 5.7 Restriction maps of *En2* YAC clones**

The three YACs are aligned relative to *En2* which is represented by the shaded boxes. Increments of 100Kb are scored on each YAC.



with a second gene. The commonality of restriction sites in CO22 and D9P8 3′ of *En2* indicates that this genomic region is faithfully represented in both clones. However an *Mlu*I site juxtaposed to the left arm of D9P8 is not present in CO22. This is likely to indicate a non-contiguous region in one of the clones. Three restriction sites to the 5′ of *En2* do not correspond between D9P8 and CO22. A *BssH*II site 275kb upstream of *En2* in C022 is absent from D9P8, while *Mlu*I and *BssH*II sites 260kb and 375kb upstream of *En2* respectively in D9P8 are absent from C022. This indicates a deletion, rearrangement or chimerism in one or both clones. YAC 98D8 contains the *En2* locus and approximately 90kb of 3′ flanking sequence. Only a *BssH*II site is present at the 5′ of *En2*, the other restriction sites constituting the CpG island observed in C022 and D9P8 being absent. It is orientated by the *Sal*I site 50kb downstream of *En2*.

To assess for chimerism D9P8 was analysed by FISH to metaphase chromosomes, as discussed in chapter 3, 3.2.5. 98D8 and C022 were not analysed as they were isolated at a later date. YAC DNA was prepared to use as a probe by catch-linker PCR (Shibasaki *et al.*, 1995) and then used by Muriel Lee for FISH analysis according to Fantes *et al.* (1995). Figure 5.8 shows representative results observed in 12 metaphase spreads. The D9P8 probe produces a unique signal, arrowed, which is located in the centromeric region of chromosome five. This is in concordance with the genetic location of *En2* (Joyner and Martin, 1987, Martin *et al.*, 1990). As no other signal is seen at any other chromosomal location D9P8 was reasoned to be a non-chimeric YAC.

The isolation of end clone fragments from each YAC was attempted using the inverse PCR protocol described by Arveiler and Porteous, (1991), as discussed in chapter 3, 3.2.4. Of the six possible end clones three were isolated, one from each YAC, as shown in table 5.4. To generate versatile markers from these end clones, sequence was generated from each fragment such that PCR primers could be designed which would amplify from genomic DNA. The use of nested PCR to generate an end

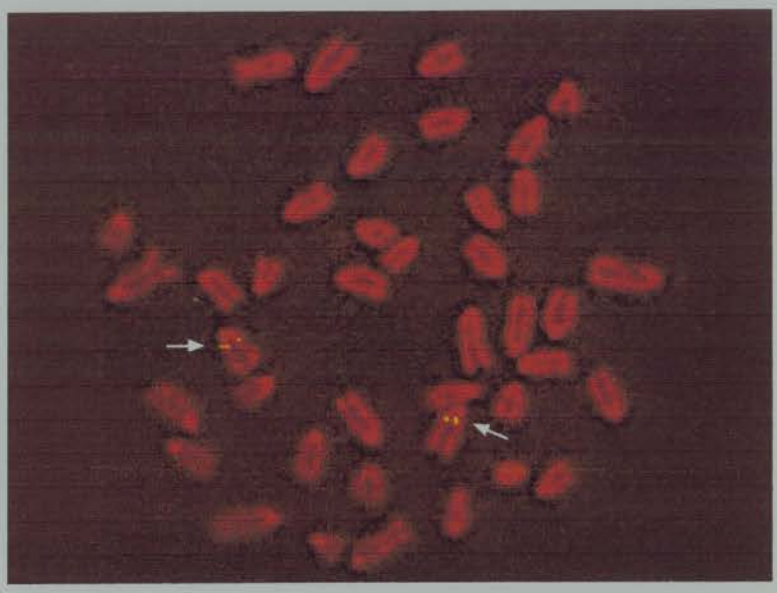


**Figure 5.8 Fluorescent in Situ hybridisation of YAC D9P8 to metaphase chromosome spreads**.

The D9P8 probe is FITC labelled against propidium iodide stained chromosomes. The FITC signal, arrowed, is located uniquely to the centromeric region of chromosome 5.

clone and the ability to read through YAC vector sequence into genomic sequence confirmed the origin of the amplified products as being the terminal sequences from YAC inserts. PCR amplification from YAC template DNA using the end clone PCR primers also confirmed their origin.

**Table 5.4 End clones isolated from *En2* YAC clones**

| YAC<br>RE = right hand end<br>LE = left hand end | Size of end clone | PCR primers designed | YACs containing end clone |
|---|---|---|---|
| D9P8RE | 650bp | D965 / D966 | D9P8 |
| C022RE | 940bp | G807 / G808 | C022 |
| 98D8LE | 300bp | F920 / F921 | 98D8, D9P8, C022 |

Despite repeated attempts the other three end clones could not be isolated, perhaps due to inconveniently placed restriction sites, as suggested by Arveiler and Porteous, (1991). The vectorette method of Riley *et al.*, (1990) was also attempted but was unsuccessful in generating additional end clones.

From the aligned restriction maps of the YACs one would predict D9P8RE, to be present in C022 if both YACs are contiguous. However PCR analysis and subsequently hybridisation of D9P8RE to Southern blots of C022 DNA demonstrated the marker was not present in C022 (data not shown). This further indicates D9P8 and/or C022 to be non-contiguous as suggested by the different restriction patterns seen upstream of *En2*. To determine the genetic location of D9P8RE and C022RE polymorphisms were sought between C57BL/6, DBA and *Mus spretus* using the rationale described in chapter 3 (3.2.5). No polymorphism was detected for C022RE but a size polymorphism between C57BL/6 and *Mus spretus* was observed for D9P8. Amplification with PCR primers D965/D966 from C57BL/6 produced a fragment of ~570bp, compared to a fragment of ~610bp from *Mus*

*spretus*. This enabled the locus to be mapped on the BSS backcross, an interspecific backcross panel of 94 mice [(C57BL/6J × *Mus spretus*)F1 × *Mus spretus*] that had previously been characterised for over 450 genetic markers throughout the genome (Rowe *et al.*, 1994). An SDP was derived by PCR amplification for a subset of 20 mice which had been typed for the proximal chromosome 5 markers *D5Mit1*, *En2* and *Mpmv-7*. *D1Mit1* maps 10cM proximal to *En2* while *Mpmv-7* maps 14cM distal to *En2*. D9P8RE should therefore map between these two markers and show concordance with *En2*. As illustrated in Table 5.5 the SDP is however non-concordant with this chromosomal region. Although FISH analysis had indicated that D9P8 was a contiguous clone this analysis demonstrates that D9P8 is in fact chimeric, the right hand end of D9P8 not being from the proximal region of chromosome 5. That this chimerism was not detected by FISH indicates a small chimeric region of less than 20kb (J. Fantes, pers. comm.).

### 5.2.6 Screening of YAC clones for novel *Gli* genes

To determine the presence of a novel *Gli* gene in D9P8 or C022 reduced stringency hybridisation was performed with a *Gli2* zinc finger probe, generated by PCR amplification of a 583bp fragment from the *Gli2* cDNA using primers D350/D514. Total yeast DNA was isolated from D9P8 and C022. 5μg of DNA was digested with *BamH*I, *EcoR*I, *Hind*III, *Pst*I and *Xho*I, separated on a 1% agarose gel and transfered to a nylon membrane. DNAs from C57BL/6 and AB1380, the YAC host yeast strain, were run on the same gel to act as positive and negative controls. Hybridisation was performed at low stringency: 5 x SSPE, 5 x Denhardts' solution, 0.5% SDS at 50°C for 16 hours. Washing was performed at 50°C in 1 x SSPE, 0.1% SDS for 20 minutes followed by 10 minutes at 50°C in 0.5 x SSPE, 0.1% SDS.

The result of this experiment, for D9P8 is shown in figure 5.9. Multiple bands are seen in each lane of mouse genomic DNA. The *Gli2* probe was generated from cDNA in a region which spans two introns (Ruppert *et al.*, 1988). The

**Table 5.5 Strain distribution pattern for D9P8RE on BSS backcross progeny**

The haplotype for each animal at each locus is indicated; C57BL/6 B, *Mus spretus* S. Recombination between markers is indicated by ×. The haplotypes for D9P8RE nonconcordant with *En2* are underlined, while those haplotypes nonconcordant with the region are double underlined.

| Locus | BSS backcross progeny | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1D | 1G | 2D | 2F | 3F | 4H | 5F | 5H | 6E | 8C | 8E | 8F | 8H | 9B | 9D | 10 | 10 | 11 | 11 | 12 |
| *D5Mit1* | B | S | B | B | B | B | S | S | B | B | B | B | S | B | S | B | S | B | B | S |
| *En2* | B | S | B | B | B | B | S | S | × S | B | B | × S | S | B | S | B | S | B | B | × B |
| *Mpmv7* | B | S | × S | × S | B | × S | × B | S | S | × S | B | S | × B | B | S | B | S | B | B | B |
| *D9P8RE* | B | B | B | B | B | S | B | B | S | B | B | B | B | B | S | S | S | B | S | B |

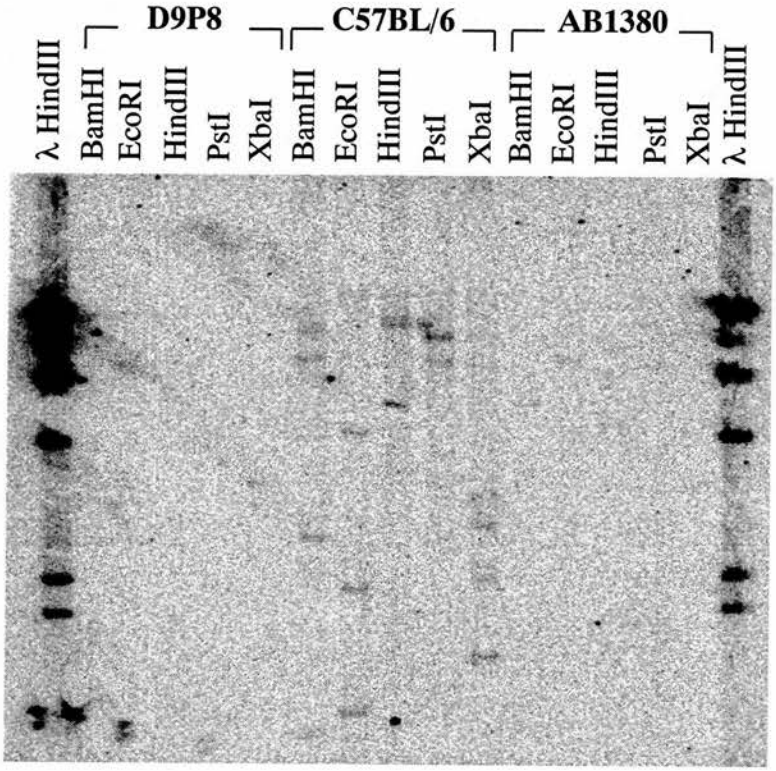



**Figure 5.9 Reduced stringency hybridisation of *Gli2* to YAC D9P8**

*Gli2* zinc finger region (PCR amplified using primers D350/D514) hybridised to YAC D9P8 at low stringency. No hybridisation to D9P8 can be seen.

restriction pattern for *Gli2* alone was not determined therefore multiple bands in each lane may represent *Gli2* if the relevant restriction sites are present in the intronic sequence spanned by the *Gli2* probe. These are likely to be the most intense bands. However several less intense bands can be seen, for example a 4.5kb band in the *EcoR*I lane or 7kb band in the *Pst*I lane. These could represent *Gli2* fragments which are partially digested or which contain small regions of probe homology. However they are likely to represent either *Gli* or *Gli3*, indicating that under these conditions the probe is hybridising to related genes. Ruppert *et al.* (1988) observe such cross-hybridisation using similar conditions.

As a novel *Gli* gene at the *Hx* locus is hypothesised to be most homologous to *Gli2* these conditions should detect such a gene in the YACs analysed, should one be present. However no hybridisation in D9P8 can be detected. A similar result was obtained for C022, indicating that no *Gli*-like genes are present in these YAC clones.

The three YACs were also used as template DNA in PCR reactions using primers E187/E197 which are degenerate for mammalian *Gli* genes (see section 5.2.3 above). No fragment could be amplified, again indicating the absence of a novel *Gli* gene close to *En2*.

## 5.3 Discussion

The elucidation of paralogous genetic regions in mammalian genomes provides a novel approach to gene identification. This strategy, termed paralogy mapping by Katsanis *et al.* (1996) exploits the existence of paralogous gene family members in chromosomal regions which have arisen following ancient duplication events. In essence, the map position of a gene within a chromosomal region which is paralogous with a second chromosomal region, predicts the existence of a related gene within this second region. For example, analysis of gene sequence and mapping data by Katsanis *et al.* enabled the discovery of a triplicated region in the human genome on chromosomes 1, 6 and 9, involving 19 gene families. Orthologous loci show similar association in the conserved syntenic regions of the mouse genome. Paralogous *PBX* loci were observed on chromosomes 1 and 9 while paralogous *NOTCH* loci were present on chromosomes 6 and 9. This lead to the prediction of novel *PBX* and *NOTCH* genes on chromosomes 6 and 1 respectively, genes which were subsequently cloned. Indeed the map location of the putative *NOTCH* gene enabled a directed cloning approach from a flow-sorted chromosome 1 cosmid library.

This predictive approach should be applied with caution due to the mechanisms which can obscure the direct relationship of relative map position between paralogous loci. Following duplication, regions may be fragmented by subsequent rearrangements or genes may be lost by deletion or through the accumulation of mutation due to the absence of selective evolutionary pressure (Lundin, 1993). However as gene sequence and mapping data accumulate regions of paralogy are becoming increasingly defined. This will enable paralogy mapping to be applied with increasing confidence to the identification of new gene family members and uncloned disease loci.

Tight genetic linkage of *Dh* and *Hx* relative to the paralogous linkage groups defined by the *En*, *Gbx* and *Htr5* genes on chromosomes 1 and 5 suggest that the two mutant alleles reside within these regions of paralogy. Given the phenotypic similarities between *Dh* and *Hx*, as detailed in chapter 1 (1.3) and the map position of each gene it

each gene it is provocative to suggest that these mutations also represent paralogous genes in this region. Should this hypothesis be correct then identification of the gene for either mutation will concomitantly suggest the existence of a paralogous gene responsible for the second mutation. At the time of this study *Gli2* was identified as a strong candidate for *Dh*, thus suggesting the presence of a *Gli2* paralogue on chromosome 5 which may be involved in *Hx*. Therefore to identify such a gene, degenerate PCR technology was employed. In addition this approach would identify additional *Gli* genes elsewhere in the genome which would be of considerable interest due to the developmental roles known to be played by the *Gli* gene family.

Amplification from 12.5 day total RNA using primers degenerate to the *Gli* specific region of the zinc finger motif specifically identified *Gli*, *Gli2* and *Gli3* but failed to detect a related gene. The bands produced by the fingerprinting procedure, figure 5.4, which represent the three *Gli* genes are at approximately equal intensity. Also the sequenced clones, which represent a random coverage of the bands, identified *Gli*, *Gli2* and *Gli3* approximately the same number of times. This indicates that at this developmental stage, in the whole embryo, the three genes are expressed at approximately equal levels. This is in agreement with the findings of Hui *et al.* (1994) who, by northern blot analysis, demonstrate similar expression levels of the three *Gli* genes at day 12.5.

Two caveats may have prevented the successful amplification of a novel *Gli* gene using the fingerprinting procedure. Firstly a novel *Gli* transcript may be absent or represented at a low level in any given RNA preparation. Secondly a novel gene may display sequence divergence at the primer binding sites which prevents successful amplification. The degenerate PCR procedure was therefore repeated using a second pair of primers which amplify within a single exon of the zinc finger region and can therefore be used to amplify from genomic DNA. When the PCR was performed at moderate stringency, at an annealing temperature of 45°C, the reaction was *Gli* specific, as no artifactual fragments were detected, however neither *Gli3*, nor a novel *Gli* gene was represented. If selective bias is seen between the amplification

of *Gli*, *Gli2* and *Gli3* then the amplification of a related gene may be prevented. The PCR was therefore repeated at a lower stringency, at an annealing temperature of 37°C for the first five rounds of the reaction. This resulted in the equal amplification of *Gli*, *Gli2* and *Gli3*, however again no additional *Gli* gene was detected.

This data strongly suggests that the mouse genome does not contain additional *Gli* family members. A novel *Gli* gene in the *Hx/Hm* region on proximal chromosome 5 is predicted to be a paralogue of *Gli2*. Such a gene would therefore be expected to amplify under the conditions employed. *Gli*-like genes may however exist which were not detected because sequence divergence from the consensus sequence of the *Gli* zinc finger region prevented amplification by the degenerate primers used. However because two different sets of PCR primers both delineated the *Gli* family to three members this is unlikely. Ruppert *et al.* (1988) have employed a *GLI* cDNA probe under low stringency hybridisation conditions to isolate related sequences from the human genome. They isolated *GLI*, *GLI2* and *GLI3* but no other *GLI* gene. However they also isolated four further zinc finger genes, termed human Krüppel related 1-4, *HKR1-4*, which show similarity to *GLI* only in Krüppel consensus regions of the zinc fingers. That they identified the three known *GLI* genes and less related genes in this screen also suggests that only three mammalian *GLI* genes exist.

A second approach was undertaken in an attempt to identify candidate genes for *Hx* and *Hm*. Previous experimental data indicated that *En1* and *Gli2*, the closest flanking markers to *Dh*, were physically linked, by a maximal distance of 600kb. The *Dh* gene, be it *Gli2* or an uncharacterised gene would therefore be defined to lie within this physical region. The paralogy between the *Dh* and *Hx* regions therefore suggests that a similar physical distance may separate *En2* from a novel *Gli* gene and that the *Hx* mutation, if not within the novel *Gli* gene, would lie in this interval. Cloning of this region would therefore provide the molecular resource to characterise the *En2* - *Hx* interval and identify candidate genes including the putative *Gli2* paralogue.

Three YAC clones containing *En2* were isolated. The contig extends 600kb to the 5′ and 300kb to the 3′ of *En2*. The two clones which span this distance were analysed for the presence of a *Gli* gene by hybridisation to a *Gli2* probe at low stringency but no signal could be detected. In addition PCR amplification using the degenerate *Gli* primers E187/E197 failed to detect a *Gli* gene indicating that these two clones do not contain a *Gli* gene. However with the exception of 200kb to the 3′ of *En2* the contig is uncorroborated. This is an important factor due to the prevalence of chimerism and rearrangement within YAC clones which can disrupt the faithful representation of the genome within the contig. The right arm of YAC D9P8 is shown to be chimeric by placement of the marker D9P8LE on an interspecific backcross, however because FISH analysis failed to detect chimerism in this YAC the extent of the chimeric region is likely to be less than 20kb. This suggests chimerism or rearrangement to the 5′ of *En2* in YAC CO22 due to a *BssH*II site located 290kb upstream of *En2* which is not present in D9P8. 220kb downstream of *En2*, an *Mlu*I site juxtaposed to the left arm of D9P8 is absent from CO22 which also indicates a region of chimerism or rearrangement in D9P8 or CO22.

Confirmation and extension of the YAC contig would be necessary to continue the analysis of this region. This would require the isolation of additional YAC clones, using *En2* and end clone markers. The orientation of the contig relative to the chromosome needs to be established to enable the contig to be extended distal to *En2*, towards *Hx* and *Hm*. This was not possible as an informative polymorphism could not be generated from the contig which would enable the genetic ordering of markers. Additional YAC clones may provide such markers.

The premise for this strategy is based on the physical linkage of *En1* and *Gli2* to the same 600kb *BssH*II and *Mlu*I fragments, as determined by PFGE analysis. However subsequent data have demonstrated this linkage to be incorrect. As detailed in chapter 3, the genomic map and the YAC contig for the *Dh* region fail to demonstrate a physical linkage between *En1* and *Gli2*. Indeed the two loci are a minimum of 1.45Mb apart. The *BssH*II and *Mlu*I fragments of approximately 600kb

which hybridise to *Gli2* and *En1* can be seen in figure 3.17 (chapter 3). However the resolution of this gel is greater than that previously achieved and alignment of the phosphorimages enables these fragments to be distinguished. No prediction can therefore be made about the physical linkage of a putative *Gli2* paralogue linked to *En2*. The strategy of cloning the genomic region distal to *En2* inorder to isolate a novel *Gli* gene therefore becomes non-viable. In addition the degenerate PCR analysis strongly suggests that the mouse genome contains only three *Gli* genes. The *En2* YAC contig could be extended towards *Hx/Hm* and this molecular resource used to isolate additional genes which could be assessed for *Hx/Hm* candidacy. However the genetic distance between *En2* and *Hx/Hm* is 1.1cM which, on average equates to greater than 2Mb. This represents a significant distance to clone through the construction of a YAC contig. Without a genetic resource to aid in the definition of the *Hx/Hm* critical region this strategy therefore becomes inefficient. For these reasons work on the *En2* YAC contig was halted.

Subsequent to this work *Inhβb* was mapped to the *Dh* critical region on chromosome 1 (see chapter 3). This information may indicate a refinement to the extent of paralogy present between chromosomes 1 and 5. *Inhβb* is located proximal to *En1* but distal to *Gli2*. One may therefore predict an *Inhβb* paralogue to map in the vicinity of *Hx/Hm* on chromosome 5. However the paralogue of *Inhβb*, *Inhβa*, is tightly linked to the *Gli2* paralogue *Gli3* on chromosome 13 (Barton *et al.*, 1989, Hui and Joyner, 1993). This suggests that the ancestral chromosome containing linked *En*, *Gli*, *Inhβb*, *Gbx* and *Htr5* genes may have undergone duplication events resulting in a triplication of this region. This seems unlikely because additional *En*, *Gbx* and *Htr5* genes have not been identified which would be predicted to map to chromosome 13. Similarly an *Inhβ* gene and a *Gli* gene would be expected to be present to chromosome 5. However the work presented here suggests the absence of additional mammalian *Gli* genes, although interestingly three novel *Inhβ* genes have recently been identified (Hötten *et al.*, 1995; Lau *et al.*, 1996; Fang *et al.*, 1996; Oda *et al.*,

1995). Although the map locations of these genes have not been determined they are less closely related to *Inhβa* and *Inhβb* than *Inhβa* and *Inhβb* are to each other. This indicates a more distant evolutionary relationship suggesting the three novel *Inhbβ* genes would not be part of these regions of paralogy.

Two scenarios are more likely. Firstly a single duplication event may have occurred followed by fragmentation in one of the duplicates resulting in the linkage groups on chromosomes 1, 5 and 13. Secondly two independent duplication events could have occurred, followed by rearrangement which juxtaposed two of the regions to form the linkage group on chromosome 1. The breakpoint for such a rearrangement appears to have occurred between the ancestral *En* and *Inhβ* genes. If *Dh* is not the result of mutation in *Gli2* or *Inhβb* then the *Dh* gene may still represent the paralogue of the *Hx* gene. A paralogous gene located distal to *Hx* and proximal to *Dh* would greatly support this prediction. If *Dh* and *Hx* are truly mutations in paralogous genes then the *Dh* gene cannot be *Gli2* or *Inhβb* but a gene located distal to *Inhβb*. If however *Dh* is the result of mutation in the *Gli2* or *Inhβb* gene then *Hx* must result from mutation in an unrelated gene.

# CHAPTER 6

# ASSESSMENT OF *Pmsc2* AS A CANDIDATE GENE FOR *Hm* AND *Hm*

## 6.1.1 Introduction

The mouse gene *Pmsc2* belongs to a superfamily of putative ATPases (Mian, 1993), members of which are involved in a diversity of cellular processes. Two functional roles have been proposed for *Pmsc2*. Firstly biochemical (Dubiel *et al.*, 1993) and genetic (Gordon *et al.*, 1993) evidence has established *Pmsc2* as a subunit of the 26S proteasome, a 2 Md complex of over 30 proteins which mediates protein degradation via the ubiquitin dependent proteolytic pathway. Secondly *Pmsc2* may modulate Tat mediated gene expression in HIV (Shibuya *et al.*, 1992).

*Pmsc2* became of interest as a potential candidate gene for *Hx* and/or *Hm* for two reasons. Firstly the gene was mapped to proximal chromosome 5 in the vicinity of *Hx* and *Hm*. Secondly expression studies showed *Pmsc2* to be differentially expressed in the developing limb bud in a manner consistent with the mutant limb phenotypes. To assess potential candidacy gene for *Hx* and/or *Hm* I have mapped *Pmsc2* both genetically and physically relative to the *Hx/Hm* loci.

## 6.1.2 The 26S proteasome and the ubiquitin dependent proteolytic pathway

The control of intracellular protein levels is crucial to defining the functional state of a cell. Schoenheimer and co-workers (1942) were the first to observe that the protein content of cells is in a constant state of flux. A delicate balance of protein synthesis and degradation, particularly of key regulatory proteins, control this dynamic state. One of the most important systems which orchestrates intracellular protein degradation is the ubiquitin-proteasome proteolytic pathway, reviewed by

Ciechanover (1994), Rubin and Finley (1995). Protein to be degraded is initially tagged with the polypeptide ubiquitin and subsequently degraded in a multi-subunit particle, the 26S proteasome. This system is present in all eukaryotes, archaebacteria and even some eubacteria (Tamura, 1995). Using protease inhibitors Rock *et al.* (1994) have shown the degradation of most intracellular protein to be mediated by the 26S proteasome. Indeed recent studies have shown ubiquitin-proteasome degradation to be instrumental in a diverse range of cellular functions such as programmed cell death, cell cycle progression, heat shock response, DNA repair, signal transduction, transcription, antigen presentation and the removal of abnormal or damaged proteins, (reviewed by Ciechanover 1994). This system is distinct from the vacuolar/lysosomal proteolytic pathways, which are primarily involved in the processing of protein imported into the cell, although they can function on cytostolic targets, especially under stressed conditions (Dice and Terlecky, 1994).

A major challenge for an intracellular proteolysis pathway is selectivity. Of the large pool of proteins within a cell specific proteins need to be targeted and destroyed at the desired time and rate. At least part of this selectivity is conferred by the tagging of proteins to be degraded via conjugation to ubiquitin, a highly stable 76 amino acid polypeptide (Ciechanover *et al.*, 1980, Hershko *et al.*, 1980, reviewed by Jentsch, 1992).

Following ubiquitination, degradation of the protein occurs in a multi-subunit protease complex, the 26S proteasome. This 2Md particle is an abundant cellular particle (up to 1% of cellular protein) and has been observed in the cytoplasm and nucleus of eukaryotes, but in no other organelles. Isolation of proteasomes yields two particles which sediment at 20S and 26S. The 20S particle represents the catalytic core of the complex which when complexed with two identical 19S particles accounts for the 26S proteasome (Hough *et al.*, 1987, Waxman *et al.*, 1987). Purified 20S particles can efficiently degrade small peptides but are inactive on intact, folded proteins (Peters, 1994). Proteolytic activity on such substrates is however restored

when the 20S particle is associated with two 19S particles. It is unclear whether the 20S particle on its own performs a functional role or whether its isolation is merely an artefact of fractionation.

Electron microscopy data (Peters *et al.*, 1993) and more recently X-ray crystallography data (Löwe *et al.*, 1995) have provided the basis for the structural model of the 20S complex. The catalytic core consists of four heptameric rings which are stacked one on top of the other to form a hollow cylinder. The proteolytic active sites are located within the middle two rings, on the inner surface of the cylinder. A crucial feature of this cylindrical structure is that the only access to the active sites is via narrow 13Å channels formed by the two outer rings. In comparison polypeptide $\alpha$-helices have a diameter of 11Å. This implies that only unfolded polypeptide is able to pass through these channels and enter the proteolytic chamber of the complex (Wenzel, 1995). This compartmentalisation provides an explanation as to why the 20S particle which is promiscuous in its protease specificity does not act randomly. Protein needs to be targeted to the opening of the proteasome, unfolded and passed through into the catalytic chamber.

The 19S particle consists of at least 16 different proteins which form a cap-like structure. The protein composition of the 19S cap is only partially known and its precise functional role remains to be firmly established. However genes for several 19S cap subunits have recently been cloned which shed light on the role of this complex. Subunit 5a, S5a, has been identified by Deveraux *et al.*, (1994) as a polyubiquitin binding protein. This suggests a mechanism whereby proteins to be degraded can dock onto the proteasome via their polyubiquitin conjugates. S5a and maybe other subunits may keep the protein tethered while it is unfolded and translocated into the proteasome. This also provides a rational as to the selectivity of the system: only ubiquitin conjugated proteins can be recognised and processed.

Other 19S subunit genes which have been recently cloned are members of a large family of ATPases (Dubiel *et al.*, 1992). Several functional roles can be

conceptually assigned to these ATPases. The assembly of the 19S cap with the 20S proteasome is ATP dependent, ATPase subunits may therefore directly interact with subunits of the 20S outer heptameric ring. The unfolding of protein may be mediated by the ATPases since the opposite process, modulation of protein folding by ATPases has been demonstrated in several systems (Hendrick and Hartl, 1993). Other energy requiring processes in which the ATPases may be involved are peptide bond cleavage, translocation of the polypeptide into the proteasome and the release of degradation products. The 19s cap complex may also play a regulatory role in terms of substrate specificity. This is discussed in section 6.1.5.

One 19S ATPase subunit recently identified is encoded by the mouse gene *Pmsc2*. Gordon *et al.* (1993) functionally cloned *Pmsc2* by its ability to rescue the conditional lethality of *mts2*, a mutant of the fission yeast *Schizosaccharomyces pombe*. *mts2* is a cell cycle mutant defective in chromosome segregation during mitosis. *mts2*$^{+}$ is the *S. pombe* homologue of human S4, which is itself an ATPase subunit of the 19S complex (Dubiel *et al.*, 1992). Relative to wild type cells, *mts2* cells accumulate high molecular weight ubiquitin conjugated protein indicating that the *mts2* phenotype arises due to defective ubiquitin dependent proteolysis. This argues that *mts2* is also a subunit of the 26S proteasome. *Pmsc2* can rescue *mts2* suggesting an interaction between the two gene products which allows *mts2* to function at the restrictive temperature.

Ghislain *et al.*, (1993) have shown that CIM5 is the *Saccharomyces ceravisiae* functional homologue of human PMSC2. *cim5-1* mutant cells arrest in mitosis also due to defective ubiquitin dependent proteolysis. This together with biochemical evidence that PMSC2 is present in the 26S proteasome (Dubiel *et al.*, 1993) strongly suggests that *Pmsc2* is a functional subunit of the 26S proteasome.

### 6.1.3 The role of *Pmsc2* as a transcriptional regulator

A further role for members of the ATPase family to which *Pmsc2* belongs has been suggested. Subunits of the 26S protease are similar or identical in their sequence to proteins which have been implicated as mediators of transcription. Shibuya *et al.* (1992) have shown similarity of the human PMSC2 gene to TBP1, 42% over 374 amino acids. TBP1 negatively regulates Tat mediated transcriptional activation in human immunodeficiency virus-1, (HIV-1). Prompted by the similarity between PMSC2 and TBP1 they investigated the involvement of PMSC2 in the regulation of HIV-1 Tat function. Using *in vitro* expression assays PMSC2 was shown to have a positive modulatory effect on Tat, possibly by regulating the interaction of Tat with the basic transcriptional machinery. Although this indicates that PMSC2 can stimulate Tat-mediated activation there is no data to suggest that this interaction occurs or is required *in vivo*.

SUG1 is another ATPase which has been shown to exhibit dual 26S proteasome and transcriptional mediation activity. Ghislain *et al.*, (1993) have shown that CIM3 (identical in sequence to SUG1) is a component of the 26S proteasome, while Kim *et al.*, (1994) and Swaffield *et al.*, (1995) have identified SUG1 in multiprotein complexes which act as mediators between the activation domains of regulatory proteins and the basic transcriptional machinery. Although the data suggests that SUG1 may play a dual role it is currently unclear as to whether *Pmsc2* is truly involved in transcription.

### 6.1.4 *Pmsc2*, a potential candidate gene for *Hx* and/or *Hm*; Chromosomal location of *Pmsc2*

*Pmsc2* was initially mapped to proximal chromosome 5 by Cathy Abbott. This was performed by analysis of the BSS backcross (the Jackson Laboratory Interspecific Backcross Panel 2), a panel of 94 mice [(C57BL/6J × *Spret/Ei*)F1 × *Spret/Ei*] that had previously been characterised for over 450 genetic markers throughout the genome (Rowe *et al.*, 1994). In order to follow *Pmsc2* in this cross an

informative polymorphism between C57BL/6J and *Mus spretus* was determined. PCR amplification across an intron within the *Pmsc2* gene using primers F1.1 and R5.5 produced a fragment of approximately 1.6 kb from C57BL/6J and approximately 1.2 kb from *Mus spretus*. This enabled a strain distribution pattern (SDP) for *Pmsc2* to be generated based on the presence or absence of a C57BL/6J allele in each of the backcross progeny, F1 mice having been backcrossed to *Mus spretus*.

Analysis of this SDP, using Map Manager (Manly 1993), indicated that *Pmsc2* was fully concordant with *Xmv45*, *Htr5a* and *Pep1b*, loci previously mapped to proximal chromosome 5 (Frankel *et al.*, 1989; Frankel *et al.*, 1992; Danielson *et al.*, 1994; Matthes *et al.*, 1992). The closest flanking markers were *D5Mit1* proximally and *Mpmv7* distally (figure 6.4). According to the consensus map for chromosome 5 (Kozak and Stephenson, 1995) this placed *Pmsc2* in a 26 cM region containing several developmental mutations including *Hx* and *Hm* (figure 6.4).

### 6.1.5 Developmental expression of *Pmsc2*

Expression data, produced by Laurence Colleaux, suggests that *Pmsc2* plays a role during development (Allen, *et al.*, 1997). Northern blot analysis shows an increase in expression from E9.5 until E14.5 followed by a decrease at E16.5. Spatial distribution of *Pmsc2* during development was investigated by whole mount *in situ* hybridisation. Early stage embryos (E8.5) show a low and diffuse signal. At E11.5, the *Pmsc2* gene is highly expressed at some specific sites. In both forelimb and hindlimb buds, the *Pmsc2* gene is expressed in a distally restricted fashion with the highest levels in the mesenchymal cells at the rim of the bud under the apical ectodermal ridge (AER). No strong hybridisation signal is present in the AER itself. This site of expression may therefore correspond to the progress zone. In E12.5 embryos *Pmsc2* is expressed in both fore and hindlimbs, in the mesenchyme lining the digits but not throughout the whole interdigital region.

Another area of *Pmsc2* expression is in the developing face. In E11.5 embryos a strong signal is detected in the mandibular and the maxillary components of the first branchial arch, the lateral and the medial nasal processes and in the auricular hillocks of the first branchial cleft associated with the formation of the external ear. At E12.5, a high level of expression is visible in lower and upper lips and in the mesenchymal cells of the nasal process. At E11.5, high levels of expression are also found in the posterior neural tube and the presomitic and somitic tissues at the tip of the tail. Finally, a strong signal is found in the genital ridge.

This expression pattern can be postulated to be consistent with the phenotypes of both *Hx* and *Hm*. As fully described in chapter 1 (1.3.4), a characteristic of both these mutants is aberrant programmed cell death (PCD) during limb morphogenesis. In *Hx* heterozygotes Knudsen and Kochhar (1981) have described alterations in the normal pattern of PCD in the developing limb bud. The opaque patch is a zone of mesodermal cells which normally undergo PCD at stage 19-20 resulting in the separation of tibial and fibular precartillage rudiments. In the mutant this zone is extended to the preaxial side resulting in a degeneration of the distal portion of the pre-cartilaginous tibia. This defect is thought to lead to tibial hemimelia. The normal pattern of PCD is also disrupted in the preaxial portion of the AER and in the underlying mesoderm of the preaxial progress zone. Through stages 17-20 preaxial AER appears in an unusually thickened state with the basal cell layer remaining abnormally proliferative and failing to undergo any normal PCD. The preaxial mesoderm of the progress zone similarly displays a greatly reduced level of characteristic PCD. It is likely that the lack of PCD in the preaxial mesoderm results in a large population of cells which remain under the prolonged trophic influence of the AER. Proliferation of these cells then produces the characteristic preaxial protrusion from which supernumerary digits form.

Cells showing the characteristic morphology of normal PCD are seen in *Hx* mice, the mutation is therefore unlikely to affect the cell death mechanism itself. Rather the mutation results in a misplaced pattern of cell death. The mechanism

which drives this aberrant PCD pattern is unknown but could be due to a defect in the normal pattern formation process itself or the mutation could affect a process, a secondary effect of which produces the abnormal pattern of cell death. The expression pattern of *Pmsc2* in the limb overlaps spatially and temporally with the first observable manifestations of the mutant phenotype; preaxial outgrowth of the limb bud at stage 17, equivalent to E11.5. This indicates the potential for *Pmsc2* to be responsible for the *Hx* phenotype.

In *Hm* embryos there is a failure of normal PCD in a restricted region of the limb bud, in the interdigital mesenchyme between digits two and five. PCD however progresses normally in the interdigital mesenchyme between digits one and two and in the anterior and posterior necrotic zones (Zakeri *et al.*, 1994). As with *Hx* this implicates a defect controlling the pattern of PCD within the limb bud, rather than the PCD mechanism itself. During normal limb development PCD in the interdigital mesenchyme is first observed at E12.5 (Zakeri et al., 1994). Since Pmsc2 is expressed in the margins of the interdigital mesenchyme at E12.5, a possible involvement in the PCD process and therefore the *Hm* phenotype is suggested.

Further to the expression pattern of *Pmsc2* suggesting a role in either or both of these mutations a recent report by Dawson *et al.* (1995) provides evidence for the involvement of the 26S proteasome in the regulation of PCD. They have looked at the expression of 26S proteasome genes in the intersegmental abdominal muscles of the tobacco horn moth *Manduca sexta*, a proportion of which undergo characteristic PCD at eclosion. The expression of a 30 kD 20S proteasome subunit showed a 5 fold increase in expression in the pre-eclosion period, just prior to the onset of PCD. A similar increase in 26S protease activity was observed. Expression of *PMSC2* and *MS73*, a novel 26S ATPase, also increased 5 fold, but did not increase relative to the 20S subunit. This upregulation of the 20S protease suggests it plays an important role during PCD. In support of this a massive increase in the expression of a polyubiquitin gene and the accumulation of ubiquitinated proteins has been seen in these cells just prior to PCD (Schwartz *et al.*, 1990a, Schwartz *et al.*, 1990b).

However most interestingly they demonstrated a differential increase in the expression of two 26S ATPases; *MTS2*, increased 3 fold and *TBP1* decreased 5 fold relative to the 20S subunit.

A possible explanation is that new proteasomes are being produced, the 19S cap complexes of which have been reprogrammed in response to the new developmental situation. The subunits of the 19S cap may confer substrate specificity to the proteasome, enabling rapid degradation of only the subset of ubiquinated protein necessary to achieve the desired change in cellular state. The role of *Pmsc2* in the developing vertebrate limb may thus be to modulate substrate specificity for the 26S proteasome in a manner which affects, either directly or indirectly, the regulation of PCD. The aberrant pattern of PCD seen in *Hx* and *Hm* may therefore be produced by mutation at the *Pmsc2* locus.

## 6.2 Results

### 6.2.1 Assessment of the genetic position of *Pmsc2* relative to *Hm* and *Hx*; Assignment of *Hx/Hm* flanking markers on the BSS backcross

A fundamental criterion for establishing candidacy of a gene for a mutation is that the gene maps within the genetically defined region of the mutation. I therefore sought to refine the genetic map location of *Pmsc2* such that its position relative to the *Hm/Hx* loci could be established. A cross in which *Hx* or *Hm* was segregating was not available for analysis, therefore linkage of *Pmsc2* could not be tested directly. However Martin and co-workers (1990) have determined *Hx* to be flanked by *En2* proximally, (1.1 ± 1.1 cM) and by *Emv1*, (2.1 ± 1.5 cM) distally. The position of *Pmsc2* relative to *Hx* and *Hm* can hence be indirectly assessed by ascertaining the position of these flanking markers relative to *Pmsc2*.

Informative markers were not initially available for placement of these two loci on the BSS backcross. To determine a polymorphism at the *En2* locus PCR primers were designed in the 3' untranslated region of *En2* (Logan *et al.*, 1992). F886 and F887 amplified across a TA dinucleotide repeat of 22 copies to give a fragment of 232 bp. No size difference was detected between C57BL/6J and *Mus spretus*, however restriction analysis identified a polymorphic *Mae*II site. This site was absent in C57BL/6J but present in *Mus spretus* to give 189 bp and 43 bp fragments. To avoid typing each of the 94 animals of the BSS backcross a subset of 20 mice, which includes the 9 animals shown to be recombinant between *D5Mit1* and *Mpmv7*, were chosen for analysis. *Mae*II digestion of F886/F887 amplified product from this panel of informative recombinants is shown in figure 6.1. The deduced SDP from this analysis is fully concordant with the SDP obtained for *Pmsc2* (table 6.1).

*Emv1* was not placed on this cross because at the time of analysis no informative marker was available. However the panel of informative recombinants was typed for two microsatellite markers, *D5Mit3* and *D5Mit66*, which according to

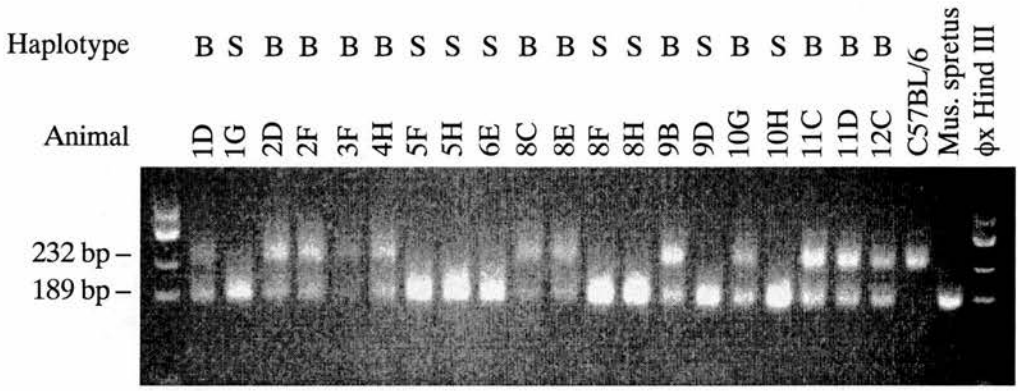



**Figure 6.1 Haplotype analysis of the BSS backcross for *En2***

*Mae*II digestion of the 232bp PCR product amplified using *En2* primers F886/F887 generates 189bp and 43bp fragments from the *Mus spretus*, while C57BL/6 does not contain a *Mae*II site. Because the F1 mice were backcrossed to *Mus spretus*, the presence of a 232 bp band indicates a C57BL/6J haplotype, B, whereas absence indicates a *Mus spretus* haplotype, S. The hapoltype of each animal is shown.

**Table 6.1 Strain distribution pattern for markers placed on the BSS backcross**

The haplotype for each animal at each locus is indicated; C57BL/6 B, *Mus spretus* S. Recombination between markers is indicated by ×.

| Locus | BSS backcross progeny | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1D | 1G | 2D | 2F | 3F | 4H | 5F | 5H | 6E | 8C | 8E | 8F | 8H | 9B | 9D | 10G | 10H | 11C | 11D | 12C |
| *D5Mit1* | B | S | B | B | B | B | S | S | B | B | B | B | S | B | S | B | S | B | B | S |
| | | | | | | | | | × | | | × | | | | | | | | × |
| *D5Mit72* | B | S | B | B | B | B | S | S | S | B | B | S | S | B | S | B | S | B | B | - |
| *D5Mit73* | B | S | B | B | B | B | S | S | S | B | B | S | S | B | S | B | S | B | B | - |
| *Xmv45* | B | S | B | B | B | B | S | S | S | B | B | S | S | B | S | B | S | B | B | B |
| *Pmsc2* | B | S | B | B | B | B | S | S | S | B | B | S | S | B | S | B | S | B | B | B |
| *En2* | B | S | B | B | B | B | S | S | S | B | B | S | S | B | S | B | S | B | B | B |
| | | | | | | | × | | | | | | | | | | | | | |
| *D5Mit3* | B | S | B | B | B | B | B | S | S | B | B | S | S | B | S | B | S | B | B | B |
| *D5Mit66* | B | S | B | B | B | B | B | S | S | B | B | S | S | B | S | B | S | B | B | B |
| | | | × | × | | × | | | | × | | | × | | | | | | | |
| *Mpmv7* | B | S | S | S | B | S | B | S | S | S | B | S | B | B | S | B | S | B | B | B |

the consensus map of chromosome 5 (Kozak and Stephenson, 1995), are located between *Hx/Hm* and *Emv1* (figure 6.4) One recombination is observed between these two markers and *Pmsc2*, placing these two markers distal to *Pmsc2*. Also placed were *D5Mit72* and *D5Mit73* which, according to the consensus map, lie either side of *Xmv45* but in this cross cosegregate with *Xmv45* and *Pmsc2* (table 6.1, figure 6.4).

From this analysis it is likely that *Pmsc2* lies proximal to *Emv1* because it is proximal to *D5Mit3* and *D5Mit66*, markers which themselves are likely to be proximal to *Emv1*. This puts *Pmsc2* on the *Hx/Hm* side of *Emv1*. *Pmsc2* cosegregates with *En1* and could therefore lie to either the distal or proximal side of *Hx/Hm*.

### 6.2.2 Location of *Pmsc2* on the European Collaborative Interspecific Backcross

To assess further the genetic position of *Pmsc2* relative to *Hx* and *Hm* a mapping study was carried out using the European Collaborative Interspecific Backcross, EUCIB. This comprises 982 progeny produced from [(C57BL/6 × *Mus spretus*)F1 × *Mus spretus*] and [(C57BL/6 × *Mus spretus*)F1 × C57BL/6] matings (EUCIB group 1994). This provides an average genetic resolution of 0.1 cM which gives the potential to resolve markers that lie an average of 200kb apart. This is an order of magnitude greater resolution than the 94 progeny of the BSS backcross, thus providing a greater probability of defining a chromosomal order for *Pmsc2*, *En2* and *Emv1*. To facilitate the fine mapping of markers within any chromosomal region 78 anchor loci, at least three on each chromosome, have been placed on all 982 progeny. This allows the identification of all the available recombinants within any interanchor interval. Panels of mice can thus be determined which will be informative for the fine mapping of a particular region. The most proximal anchor locus placed on chromosome 5, *D5Nds8*, is however distal to the *Hx/Hm* region, a panel of informative recombinants cannot therefore be deduced directly. It was therefore necessary to assign a marker that is predicted to be proximal to *Hx/Hm* and determine

those mice recombinant between it and *D5Nds8*. This panel of mice could then be typed for *Pmsc2*, *En2* and *Emv1*.

*D5Mit226* is a centromeric microsatellite marker, proximal to *En2*, and was used to type 348 mice which were chosen at random from the 982 progeny. Examples of this assay are shown in figure 6.2A. 55 of these mice showed recombination with *D5Nds8* (table 6.2).

The *Pmsc2* PCR assay, used by Cathy Abbott (primers F1.1 and R5.5), did not function in my hands. Another variant was thus determined by designing PCR primers F4 and 4R2 which amplified across an intron within the *Pmsc2* gene to give a 1kb product (*Pmsc2* cDNA sequence was obtained from C. Gordon). Restriction analysis of this amplified fragment identified a variant *Sac*I site, this site being absent in *Mus spretus* but present in C57BL/6 to give 650 bp and 350 bp fragments (figure 6.2b). The SDP obtained is illustrated in table 6.2.

An informative marker for the *Emv1* locus was determined in the following way. A 1.4 kb *Sac*I fragment was subcloned into pBluescript from the plasmid p*Emv1*, a genomic clone of the *Emv1* locus (a gift from N. Copeland). This fragment, determined to be single copy as judged by hybridisation to mouse genomic Southern blots, was used to isolate a genomic clone from a 129/Sv genomic lambda library (Stratagene). In an attempt to isolate microsatellite repeats which may be variant the clone was digested with *Sau*3AI and shotgun cloned into pBluescript. Approximately 500 colonies (a three fold redundancy) were screened by hybridisation with a $^{32}P$ end-labelled $GT_{11}$ oligonucleotide, using the method of Buluwela *et al.* (1989). Stringent washing conditions were used in order to identify uninterrupted CA repeats which are more likely to be polymorphic. Three positive clones were isolated and the plasmid inserts sequenced. In all three clones the same $CA_{22}$ repeat was identified. PCR primers which would amplify across the repeat were designed. Primers G313 and G454 amplified a 100 bp fragment from C57BL/6

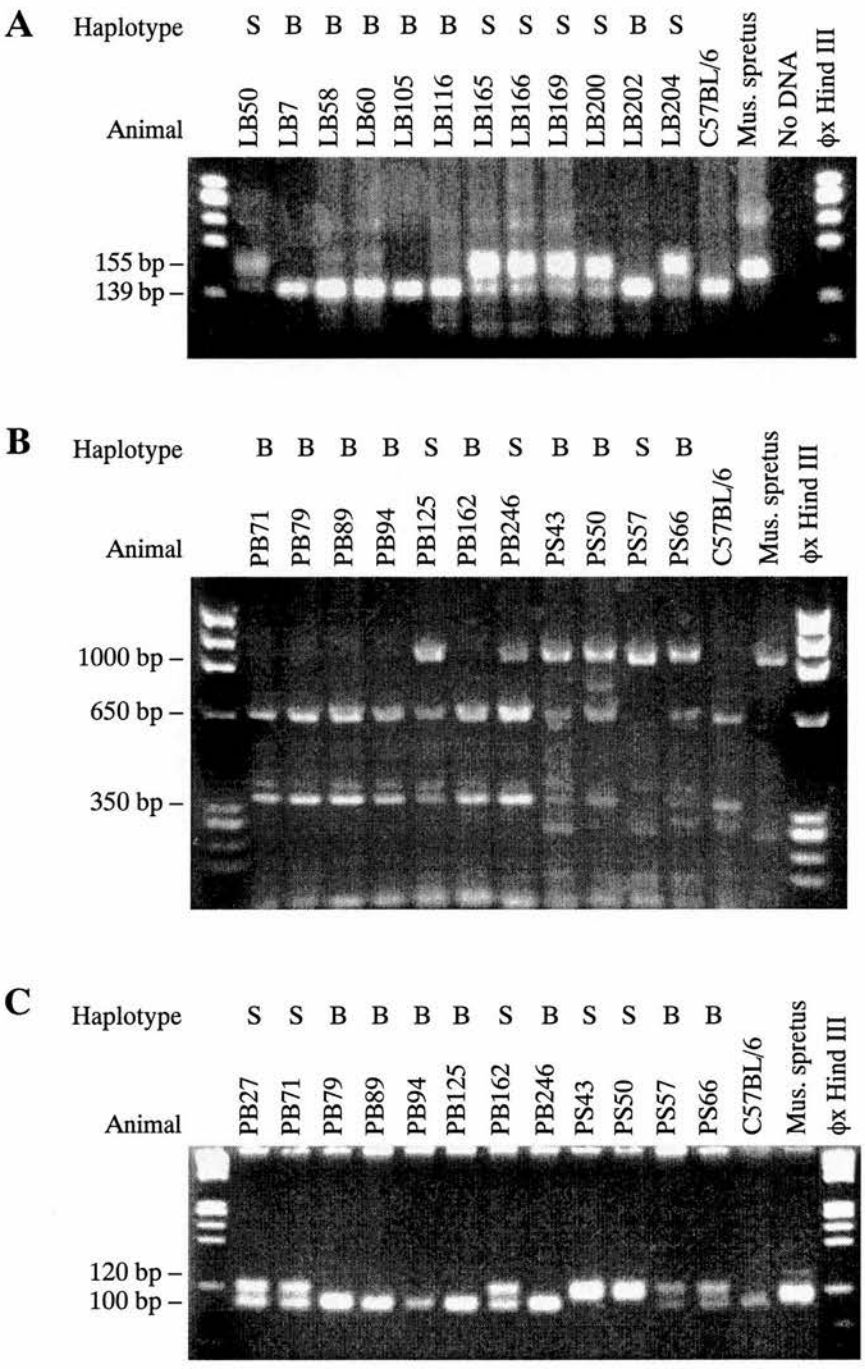

A
Haplotype
S B B B B B S S S S B S
Animal
LB50 LB7 LB58 LB60 LB105 LB116 LB165 LB166 LB169 LB200 LB202 LB204 C57BL/6 Mus. spretus No DNA φx Hind III
155 bp –
139 bp –
B
Haplotype
B B B B S B S B B S B
Animal
PB71 PB79 PB89 PB94 PB125 PB162 PB246 PS43 PS50 PS57 PS66 C57BL/6 Mus. spretus φx Hind III
1000 bp –
650 bp –
350 bp –
C
Haplotype
S S B B B B S B S S B B
Animal
PB27 PB71 PB79 PB89 PB94 PB125 PB162 PB246 PS43 PS50 PS57 PS66 C57BL/6 Mus. spretus φx Hind III
120 bp –
100 bp –

**Table 6.2 Strain distribution pattern for markers placed on EUCIB**

The haplotype for each animal at each locus is indicated; C57BL/6 B, *Mus spretus* S. Each animal is named by one of the prefixes; LS, LSF, LB, PS or PB followed by a number. Recombination is indicated by ×. Failure of a marker to type in any animal is indicated by -.

| Locus | Informative EUCIB backcross progeny | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | LS | | | | | | | | | | | | | | | | | | |
| | 41 | 48 | 54 | 70 | 73 | 78 | 81 | 87 | 95 | 120 | 198 | 217 | 221 | 223 | 263 | 269 | 317 | 319 | 325 |
| *D5Mit226* | S | S | B | B | B | S | B | B | B | B | B | B | B | S | S | B | B | S | S |
| | | | | | | | | | × | | × | × | × | × | | × | × | | × |
| *Pmsc2* | S | S | B | B | B | S | B | B | S | B | S | S | S | B | S | S | S | S | - |
| | | × | | | | | | | | × | | | | | × | | | × | |
| *En2* | S | B | B | B | B | S | B | B | S | S | S | S | S | B | B | S | S | B | - |
| *Emv1* | S | B | B | B | B | S | B | B | S | S | S | S | S | B | B | S | S | B | - |
| | × | | × | × | × | × | × | × | | | | | | | | | | | |
| *D5Nds8* | B | B | S | S | S | B | S | S | S | S | S | S | S | B | B | S | S | B | B |

| Locus | Informative EUCIB backcross progeny | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | LS | | | | | | | LSF | | | | LB | | | | | | | |
| | 337 | 375 | 584 | 718 | 719 | 722 | 726 | 829 | 830 | 853 | 872 | 58 | 139 | 207 | 239 | 304 | 437 | 441 | 442 |
| *D5Mit226* | S | B | S | B | S | S | S | S | B | B | B | B | S | B | S | B | S | S | B |
| | × | | × | | | | | | × | | | | | | | × | × | × | × |
| *Pmsc2* | - | B | B | B | S | S | S | S | S | B | B | B | S | B | S | S | - | B | S |
| | | × | | | | | × | | | × | | | | × | | | | | |
| *En2* | B | S | B | B | S | S | B | S | S | S | B | B | S | S | S | S | - | B | S |
| | | | | | | × | | × | | | × | | | | | | | | |
| *Emv1* | - | S | B | B | S | B | B | B | S | S | S | B | S | S | S | S | - | B | S |
| | | | | × | × | | | | | | | × | × | | × | | | | |
| *D5Nds8* | B | S | B | S | B | B | B | B | S | S | S | S | B | S | B | S | B | B | S |

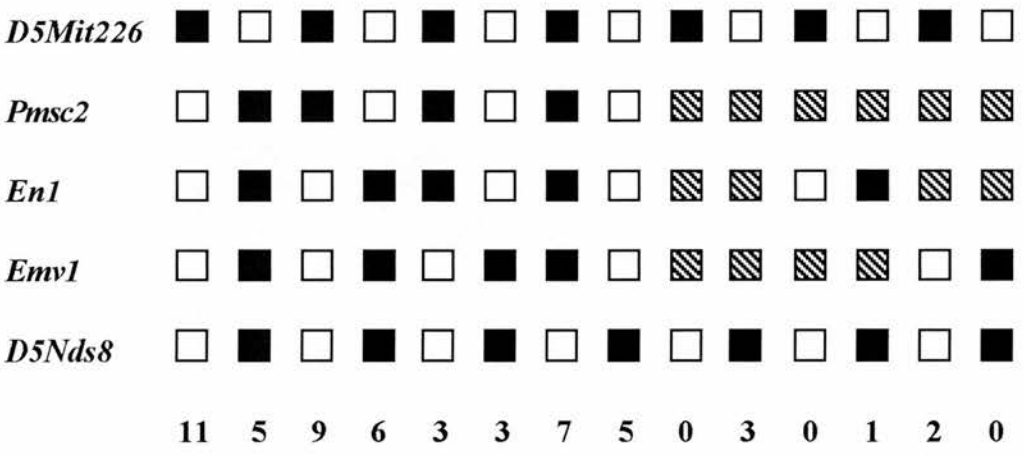
**Table 6.2(cont.) Strain distribution pattern for markers placed on the EUCIB backcross**

The haplotype for each animal at each locus is indicated; C57BL/6 B, *Mus spretus* S. Each animal is named by one of the prefixes; LS, LSF, LB, PS or PB followed by a number. Recombination between markers is indicated by ×. Failure of a marker to type in any animal is indicated by -.

| Locus | Informative EUCIB backcross progeny | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | LB | | | | | | | | | PB | | | | PS | | | |
| | 502 | 506 | 520 | 522 | 523 | 596 | 597 | 604 | 611 | 71 | 125 | 162 | 246 | 43 | 50 | 57 | 66 |
| *D5Mit226* | B | B | B | S | B | B | B | B | B | B | S | B | S | B | B | S | S |
| | | | × | × | × | × | × | × | | | | | | | | | × |
| *Pmsc2* | B | B | - | - | - | S | S | S | B | B | S | B | S | B | B | S | B |
| | | × | | | | | | | × | | | × | × | × | × | × | |
| *En2* | B | S | - | - | - | S | S | S | S | B | S | S | B | S | S | B | B |
| | × | | | | | | | | | × | × | | | | | | |
| *Emv1* | S | S | S | - | S | S | S | S | S | S | B | S | B | S | S | B | B |
| *D5Nds8* | S | S | S | B | S | S | S | S | S | S | B | S | B | S | S | B | B |

and a 120 bp fragment from *Mus spretus*. The 55 informative mice were typed, illustrated in figure 6.2c and an SDP determined (table 6.2). An SDP for the 55 mice was also determined for *En2* (table 6.2) using the polymorphism previously described. These data are summarised in figure 6.3.

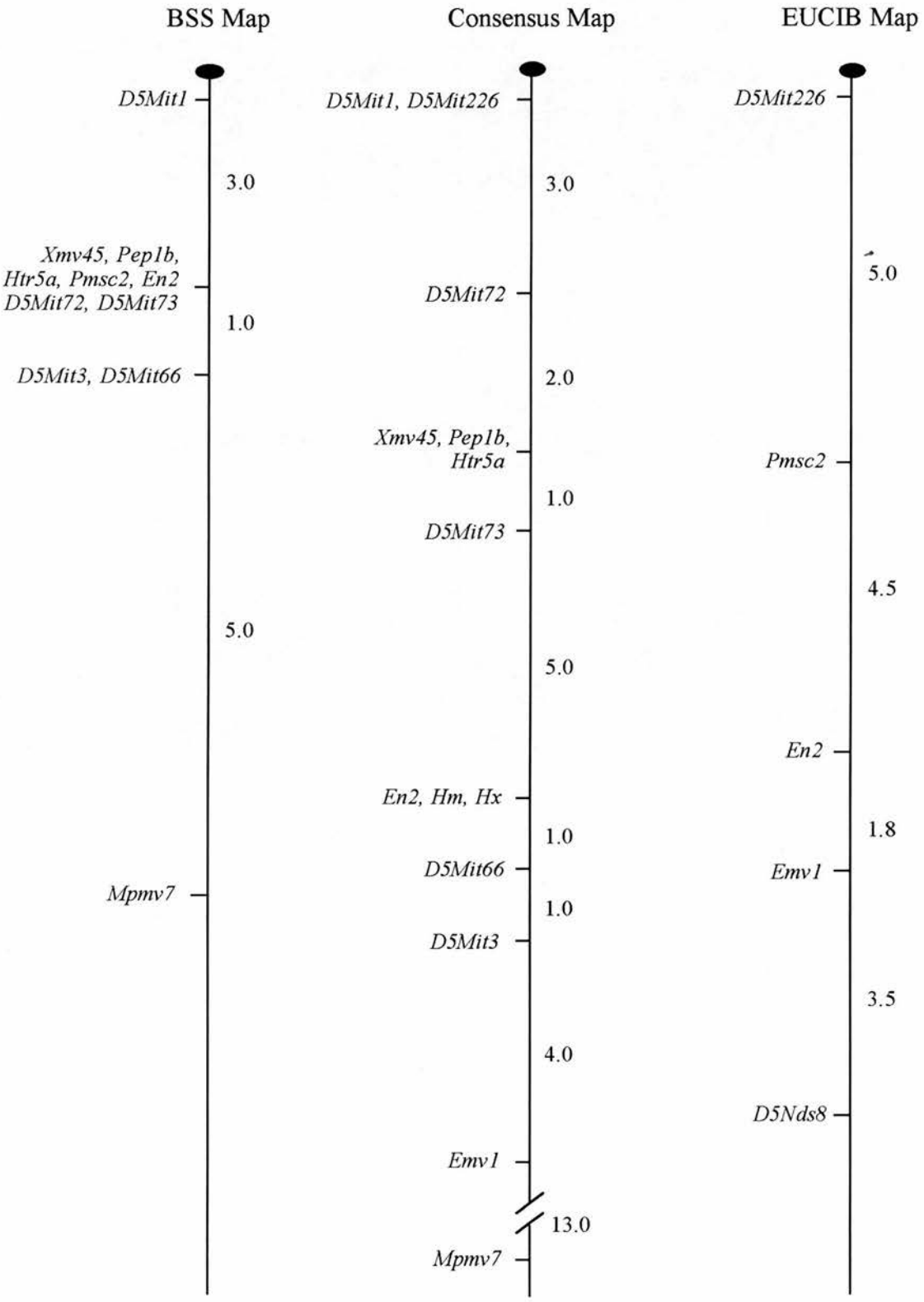
**Figure 6.3 Summary of the EUCIB analysis**
Loci mapped in the analysis are shown to the left. Each column represents the chromosome inherited from the (C57BL/6 x *Mus spretus*) F1 parent. The black boxes represent the C57BL/6 allele and the white boxes represent the *Mus spretus* allele. Chromosomes in which an allele was not defined is shown as a crosshatched box. The number of each type of chromosome identified is shown at the bottom.




The predicted gene order, based on minimising the number of double crossovers, and the relative genetic distance between loci is *D5Mit226* — 4.7 ± 1.1cM — *Pmsc2*— 4.4 ± 1.1cM — *En2* — 1.8 ± 0.7cM — *Emv1* — 3.5 ± 1.0cM — *D5Nds8*. This places *Pmsc2* proximal to *En2*, outside the genetically defined region for *Hx* and *Hm*. The genetic distances are calculated based on a total progeny number of 342, rather than the 384 progeny initially typed for *D5Mit226*. This is because six of the 55 mice recombinant between *D5Mit226* and *D5Nds8* failed to type for one or more of the *Pmsc2*, *En2* or *Emv1* markers and have

**Figure 6.4 The BSS backcross, EUCIB and Consensus Genetic maps of proximal chromosome 5**

Distances between loci, in centimorgans, are indicated on the right of each map. The consensus map is that produced by the chromosome 5 committee (Kozak and Stephenson, 1995). Maps are not drawn accurately to scale.



thus been excluded. The adjusted number of progeny is the fraction of the total 384, equal to the fraction of the mice which were successfully typed for all markers, (384 × (55 – 6) /55 = 342). Alternatively the total number of progeny typed can be adjusted by subtracting those animals which failed to type from the total analysed; 384 - 6 = 378. The recombination frequencies are then as follows -

*D5Mit226* — 4.2 ± 1.0cM — *Pmsc2*— 4.0 ± 1.0cM — *En2* — 1.5 ± 0.6cM — *Emv1* — 3.2 ± 0.9cM — *D5Nds8.*

It is possible that two recombination events could occur between *D5Mit226* and D5Nds8 which would result in both markers being of the same genotype. This would have appeared as no recombination in the initial screen and would therefore not have been taken into account. The recombination frequencies between markers may hence be slightly underestimated.

### 6.2.3 Physical mapping of *Pmsc2*, *En2* and *Emv1* by fluorescent in situ hybridisation

In an alternative attempt to localise *Pmsc2* relative to *Hm* and *Hx*, markers for *En2*, *Emv1* and *Pmsc2* were mapped to metaphase chromosome spreads by fluorescent in situ hybridisation, FISH. An application of this technology enables the simultaneous hybridisation of multiple probes which are distinguished by using fluorescent labels with different colours. Where two or more probes give discrete signals it is possible to determine a relative marker order along the chromosome. Due to the condensed nature of metaphase chromosomes the ordering of loci requires that differentially labelled markers are separated by at least 1Mb (Trask *et al.*, 1991). The genetic distance between *En2* and *Emv1*, calculated by Martin *et al* (1990), is 3.3 cM, implying a physical distance between the two of approximately 7Mb. These two loci should therefore be separable by this methodology. The ability to resolve the relative position of *Pmsc2* will depend on *En2* and *Emv1* being located at least 1Mb from *Pmsc2.*

Genomic lambda clones for each marker were isolated. The *En2* specific probe pλ4SR (Davidson *et al.*, 1988) and a *Pmsc2* 1.4 kb cDNA were used to screen a genomic 129/Sv lambda library (Stratagene). Positive clones were isolated to purity and their authenticity verified by the sequencing of PCR products amplified using *En2* or *Pmsc2* specific primers. Isolation of an *Emv1* lambda clone is already described (6.2.2). Isolated lambda DNAs were then amplified by catch linker PCR to generate probes for FISH.

The FISH experiments were performed by Muriel Lee. Determination of the relative order of the three markers was attempted by the hybridisation of pairs of probes, in all combinations, to metaphase chromosome spreads. For each pair of probes one was labelled with biotin while the other was labelled with digoxygenin. Biotin was detected with FITC and digoxygenin with Texas red which produce green and red fluorescent signals respectively.

For each probe pair distinct signals could be produced for each probe, examples of which are shown in figure 6.5. At least 20 chromosomes were analysed for each combination of probes, the results of which are shown in table 6.3. The determination of probe order was performed by Muriel Lee who, at the time of the study, was unaware of the genetic order of the three probes. The analysis of marker order was therefore performed blind.

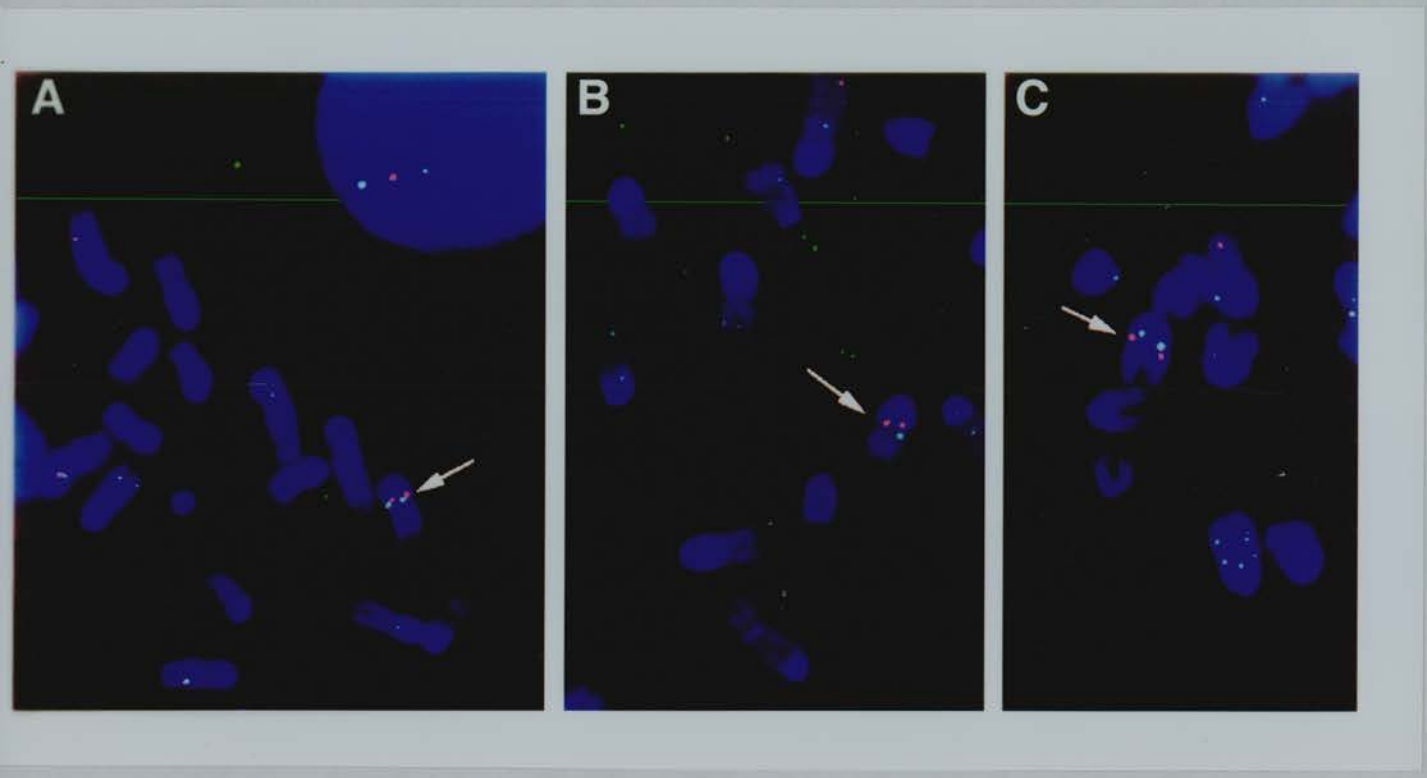



**Figure 6.5 Relative order of *Pmsc2*, *En1* and *Emv1* on chromosome 5 determined by FISH analysis**

**Panel A** shows a typical hybridisation to chromosome 5 using probes for *Emv1* (green) and *En2* (red); **Panel B** for *Emv1* (green) and *Pmsc2* (red); and **Panel C** for *En2* (red) and *Pmsc2* (green). Chromosomes are counter-stained with DAPI.

**Table 6.3 Observed frequency of marker order following FISH analysis with pairwise combinations of *Pmsc2, En2 and Emv1* probes**

| Probe combinations and marker order relative to the centromere, *Cen* | Number of chromosomes observed |
|---|---|
| ***Pmsc2 / En2*** | |
| *Cen - Pmsc2 - En2* | 20 |
| *Cen - Pmsc2 = En2* | 2 |
| *Cen - En2 - Pmsc2* | 2 |
| ***En2 / Emv1*** | |
| *Cen - En2 - Emv1* | 17 |
| *Cen - En2 = Emv1* | 2 |
| *Cen - Emv1 - En2* | 2 |
| ***Pmsc2 / Emv1*** | |
| *Cen - Pmsc2 - Emv1* | 39 |
| *Cen - Pmsc2 = Emv1* | 8 |
| *Cen - Emv1 - Pmsc2* | 1 |

In the majority of cases distinct red and green signals, corresponding to each marker, could be distinguished, one being determined to be centromeric to the other (see figure 6.5). A yellow signal was occasionally observed indicating that two probes were overlapping. Probe order could then not be distinguished. The results indicate the following marker order for each probe pair. *Pmsc2* and *En2*; *Cen - Pmsc2 - En2*, *En2* and *Emv1*; *Cen - En2 - Emv1*, *Pmsc2* and *Emv1*; *Cen - Pmsc2 - Emv1*. Combined this indicates an overall order of *Cen - Pmsc2 - En2 - Emv1*. This is in agreement with the marker order determined following analysis of the EUCIB backcross.

### 6.3 Discussion

The assessment of *Pmsc2* as a potential candidate gene for *Hx* or *Hm* was undertaken based on two observations. Firstly *Pmsc2* co-localised to the same chromosomal region as *Hx* and *Hm* on proximal chromosome five. Secondly the restricted pattern of expression in the developing limb suggested a functional role for *Pmsc2* in those areas of the limb bud which first display the mutant phenotypes. To further investigate the possibility of *Pmsc2* being responsible for either or both of these mutations I have finely mapped *Pmsc2* relative to the *Hx/Hm* loci. Both mutations have been defined to a genetic interval flanked by *En2* and *Emv1* (Martin *et al.*, 1990). Placement of *Pmsc2* relative to these flanking markers will therefore determine the gene to lie within or external to the *Hx/Hm* interval.

Initially *Pmsc2* and *En2* were mapped on the BSS backcross but no recombination was observed between them. This indicated that *Pmsc2* was tightly linked to *En2* but did not reveal the relative order of the two markers. The assignment of *Pmsc2* relative to the *Hm/Hx* region could therefore not be made. This does however suggest that *Pmsc2* is also tightly linked to *Hx/Hm* as Martin *et al.*, (1990) have mapped *En2*, the closest marker to *Hx/Hm* at the time of this study, 1.1cM proximal to *Hx* while a recent study by Robert *et al.* (1994) places *En2* 2.1cM proximal to *Hx* and *Hm*.

To resolve the map position of *Pmsc2* relative to *En2* and *Emv1* it was therefore necessary to analyse a second, larger backcross. EUCIB consists of 982 animals, enabling a genetic resolution of 0.1cM, an order of magnitude greater than the BSS backcross. Initially, to determine a panel of informative mice recombinant on proximal chromosome 5, 384 animals were typed with the centromeric marker D5Mit226. This analysis enabled the identification of 55 mice recombinant between D5Mit226 and D5Nds8, a marker previously placed on EUCIB which is known to be distal to the *Hx/Hm* interval. From this panel of 55 mice 49 were successfully typed for *Pmsc2*, *En2* and *Emv1*. As illustrated in figures 6.3 and 6.4 recombination

was observed between all three markers, the best marker order being *centromere - Pmsc2 - En2 - Emv1*. *Pmsc2* is shown to be located 4.4 ± 1.1cM proximal to *En2*. This places *Pmsc2* outside the genetically defined region for *Hx/Hm* and as such is excluded as a candidate gene for these mutations.

The ability to resolve the order of these markers on the EUCIB backcross but not on the smaller BSS backcross illustrates the value of having a mapping resource of high genetic resolution, such as EUCIB, available to the general community. Only a portion of the backcross, 384 mice, was analysed because of the substantial amount of work required to initially type all 982 mice for the centromeric marker *D5Mit226*. However the 55 informative mice selected provided the resource to determine the relative marker order. This panel may subsequently be used to quickly assess the location of any other potential candidate gene. Had the position of *Pmsc2* relative to *En2* and *Emv1* not been resolved through analysis of these mice further backcross progeny could have been analysed.

From the EUCIB analysis the distance between *En2* and *Emv1* is 1.8 cM. This result compares with that of Martin and co-workers (1990) who, using an intraspecific backcross, assign a distance of 3.3 cM. The difference in distance may reflect the relatively small number of mice, 94, analysed by Martin and colleagues. However the difference may be due to the observation that smaller recombination frequencies are usually obtained in interspecific crosses as compared to intraspecific crosses (Reeves *et al.*, 1990, Seldin *et al.*, 1989). This discrepancy is thought to result from sequence differences between *Mus* subspecies which may suppress recombination. A direct comparison of the distances calculated between the other markers placed on EUCIB cannot be made to any other backcross. This is because two or more of the markers have not been typed together on a single backcross. However the distance of 9.5 cM between *D5Mit226* and *En2* compares favourably with 11 cM estimated on the consensus genetic map, a 'best fit' map deduced from the compilation of all available mapping data, produced by the chromosome 5

committee (Kozak *et al.*, 1995). The distances calculated are therefore likely to represent comparable estimates in relation to the results generated from other crosses.

The BSS backcross failed to order *Pmsc2* relative to *En2*. This is likely to be due to a relative suppression of recombination within this region rather than a genuine tight linkage between *Pmsc2* and *En2*. This is because the two were separated by analysis of the EUCIB backcross, in a manner consistent with the consensus distances for the region. A probable threefold compression of genetic distance has occurred. The genetic distance between *D5Mit1* and *Mpmv7* deduced from the BSS backcross is 10 cM. This compares to the consensus map value of 30 cM. The *D5Mit* markers placed on the BSS backcross have also been placed on the Whitehead/MIT interspecific $F_2$ intercross (Dietrich *et al.*, 1992). The order on the BSS map is *D5Mit1* - 3.0 cM - *D5Mit72*, *D5Mit73* - 1.0 cM - *D5Mit3*, *D5Mit*66 compared to *D5Mit1* - 3.0 cM - *D5Mit72* - 3.0 cM - *D5Mit73* - 6.0 cM - *D5Mit*66 - 1.0 cM - *D5Mit3* (Whitehead/MIT, Genetic Map of the Mouse, Database release 5, 1994). The overall distance of this region is three times smaller, 4 cM versus 13 cM, in the BSS backcross. Due to the clustering of loci at the *En2*, *Pmsc2* map position this is likely to be due to a compression of the BSS map rather than an expansion of the Whitehead/MIT map.

Placement of *Pmsc2*, *En2* and *Emv1* markers on metaphase chromosomes by FISH methodology has clearly indicated their physical order along the chromosome; *centromere - Pmsc2 - En2 - Emv1*. This is in full agreement with the deduced genetic order. The ability to resolve the relative positions of the three probes indicates that they are separated from one another by at least 1Mb (Trask *et al.*, 1991). However an estimate of the approximate distance between markers cannot be made at the resolution of metaphase chromosomes. The genetic distances calculated here, *Pmsc2*— 4.4 ± 1.1cM — *En2* — 1.8 ± 0.7cM — *Emv1* — 3.5 ± 1.0cM, indicate a distance greater than 1Mb between markers.. Had two or more of the markers not

been separable on metaphase chromosomes, due to close physical proximity, FISH at a higher resolution, using interphase chromosomes could have been performed. By interphase mapping probe order has been demonstrated over distances ranging from 25kb to 2Mb (Trask *et al.*, 1989, Trask *et al.*, 1993, Yokota *et al.*, 1995, reviewed by Heiskanen *et al.*, 1996).

*Pmsc2* is a component of the 26S proteasome, a ubiquitous protease complex present in every cell. One might therefore expect *Pmsc2* to be uniformly expressed throughout the embryo. However *Pmsc2* displays a complex developmental expression pattern. The sites of expression in the limb bud correspond to areas which undergo extensive cell proliferation; in the distal mesenchyme of the progress zone at E11.5 and at the apex and along the digits at E12.5. This restricted expression suggests a possible requirement for higher levels of the 26S proteasome in these regions. Alternatively *Pmsc2* may be upregulated relative to the 26S proteasome. *Pmsc2* function may therefore be to confer specificity of action, perhaps by determining a class of substrate to be degraded. Either scenario would enable the proteolytic activity of the 26S proteasome to function in a developmentally regulated manner.

*Pmsc2* was postulated to be involved, either directly or indirectly, in the regulation of PCD due to the aberrant pattern of PCD observed in *Hx* and *Hm*. However this analysis has shown that *Pmsc2* is not responsible for either of these mutations. The developmental role of *Pmsc2* therefore remains unknown but *Pmsc2* and other proteasome ATPases may be involved in the regulation of a number of cellular systems, including PCD, which contribute to developmental fate. The involvement of these genes in development therefore warrant further study.

# CHAPTER 7

# CONCLUSIONS

The ultimate aim of the work presented in this thesis was to clone the genes responsible for *Dh* and *Hx/Hm*. For *Dh* a positional cloning strategy was undertaken, while for *Hx/Hm* investigation focused on two potential candidate genes.

The physical characterisation of the *Dh* critical region, defined genetically to a region of 1.2cM, was initiated by the generation of a long range PFGE genomic restriction map and the construction of a YAC contig. Using a combination of the closest flanking markers to *Dh*, *Gli2* and *En1*, YAC derived markers and *Inhβb*, which is non-recombinant with *Dh*, a restriction map covering 2.4Mb was generated. However no linkage was determined between *Gli2* and *En1*, either by long range restriction mapping or through overlap of *Gli2* and *En1* YAC contigs. This analysis determines the minimum size of the *Dh* region to be 1.45Mb, the map extending 1.0Mb distally from *Gli2* and 0.45Mb proximally from *En1*. *Inhβb* resides 550-600kb distal to the 3′ of *Gli2*. Completion of the map will require additional YAC walking steps using end clone markers derived from the distal end of the *Inhβb* containing YACs (119A4, 76B3, 89H7 and 20F7). Similarly the *En1* YAC contig should be extended by further walking steps, in both directions as the orientation of the contig relative to the chromosome is not known. Additional YACs may provide the resource to orientate the *En1* YAC contig, either genetically or physically if overlap with the *Gli2* contig can be established. Suitable probes, isolated from additional YACs should be placed on the long range PFGE restriction map to confirm the YAC derived map. The *Dh* region may be further delimited by the generation of markers, distal to *Inhβb*, which can be mapped on the distal or *En1* side of the three backcross animals recombinant between *Dh* and *En1*. This may determine the maximal extent of the *Dh* region without the requirement for physical linkage between *Gli2* and *En1*.

*Gli2* and *Inhβb* are shown to be non-recombinant with *Dh* and both represent good candidates for the mutation. No coding sequence or splice site mutation was detected in *Gli2* or *Inhβb* (although 40 nucleotides remain to be unequivocally sequenced from the *Dh* allele of *Inhβb*) and RT-PCR showed both genes to be expressed in *Dh* at day 12.5. The expression of both genes should be investigated further by semi-quantitative RT-PCR using RNA preparations from restricted regions at a range of developmental time points. A more detailed study of expression would however be achieved by in situ hybridisation. In addition the untranslated regions of both genes should be assessed for mutation. With failure to detect mutation in *Gli2* and *Inhβb* additional genes need to be isolated from the *Dh* region and assessed as candidates.

Continued molecular analysis will define a maximal *Dh* critical region on the basis of the informative backcross animals. Presently the minimal region which is non-recombinant with *Dh* is 550kb, between the 3′ of *Gli2* and *Inhβb*. The larger the maximal region the less likelihood of identifying the *Dh* gene. This may be addressed by the generation of an additional backcross panel in order to define a smaller genetic interval. In addition generation of an interspecific backcross panel would greatly aid the identification of polymorphic markers. More ambitiously, the *Dh* gene may be localised by complementation using large genomic clones, such as candidate YACs or BACs from the *Dh* contig, to rescue the *Dh* phenotype in transgenic *Dh* animals (e.g. see Antoch *et al.*, 1997). However if *Dh* is a hypermorph or a neomorph then complementation with the wild type allele may not be possible. If so this approach could only be employed by the use of genomic clones derived from *Dh* DNA which would be used to recreate the *Dh* phenotype on a wild type background.

Another alternative approach is to employ the comparative genomics of the pufferfish, *Fugu rubripes*. The genome of pufferfish is approximately eight times smaller than that of mammals but is estimated to contain the same number of genes, therefore gene density is assumed to be much greater (reviewed by Elgar *et al.*, 1996). In regions of conserved synteny a pufferfish genomic clone, such as a 40kb cosmid

may represent a 320kb syntenic region of the mouse genome. Identification of genic sequence in pufferfish cosmid clones has been efficiently demonstrated using a shotgun sequencing protocol. Therefore isolation of *Gli2*, *Inhβb* and *En1* cosmid clones may lead to the identification of linked genes. These genes could then be assessed for the existence of mouse orthologues in the *Dh* region.

The identification of *Gli2* as a candidate for *Dh* prompted the speculation of a novel *Gli* gene being involved in *Hx/Hm* due to the paralogous nature of the *Dh* and *Hx/Hm* regions on chromosomes one and five respectively. Degenerate PCR analysis identified the three known *Gli* genes but failed to detect a novel *Gli* gene. This strongly suggests that a novel mouse *Gli* gene does not exist. It is presently unclear if *Dh* and *Hx/Hm* are alleles of paralogous genes; additional genes mapped to these regions would help in determining whether the regions of paralogy extend to the *Dh* and *Hx/Hm* loci. Continued investigation to identify the *Hx/Hm* gene(s) would require the genetic resource of a backcross segregating for either loci, however identification of the *Dh* gene will allow the presence of a paralogue in the *Hx/Hm* region to be rapidly assessed. Conversely identification of *Hx/Hm* may lead to the identification of the *Dh* gene.

The differential expression pattern of *Pmsc2* during embryogenesis indicates a developmental role. Further the restricted pattern of expression in the developing limb suggests a functional role in areas which are affected in *Hx* and *Hm*. Localisation of *Pmsc2* to proximal chromosome five, in the region of *Hx/Hm* therefore prompted the possibility of *Hx* and/or *Hm* representing mutant alleles of *Pmsc2*. This was investigated by the placement of *Pmsc2* relative to the closest flanking markers to *Hx/Hm* both genetically, on a high resolution backcross panel, EUCIB, and physically by FISH analysis. Analysis of EUCIB determined the gene order; *Pmsc2* 4.4 ± 1.1cM - *En2* 1.8 ±0.7cM - *Emv1*. Analysis by FISH corroborated this result, indicating the same gene order. *En2* is proximal to *Hx/Hm*, therefore *Pmsc2*, being proximal to *En2* is excluded from the *Hx/Hm* region.

# BIBLIOGRAPHY


Albano, R. M., Arkell, R., Beddington, R. S. P., and Smith, J. C.(1994). Expression of inhibin subunits and follistatin during postimplantation mouse development: decidual expression of activin and expression of follistatin in primitive streak, somites and hindbrain. *Development* **120**, 803-813.

Albano, R. M., Groome, N., and Smith, J. C.(1993). Activins are expressed in preimplantation mouse embryos and in ES and EC cells and are regulated on their differentiation. *Development* **117**, 711-723.

Alcedo, J., Ayzenzon, M., Von Ohlen, T., Noll, M., and Hooper, J. E.(1996). The *Drosophila smoothened* gene encodes a seven-pass membrane protein, a putative receptor for the Hedgehog signal. *Cell* **86**, 221-232.

Alexandre, C., Jacinto, A., and Ingham, P. W.(1996). Transcriptional activation of *hedgehog* target genes in *Drosophila* is mediated directly by the Cubitus interruptus protein, a member of the GLI family of zinc finger DNA-binding proteins. *Genes & Dev.* **10**, 2003-2013.

Allen, J., Colleaux, L., Davidson, D., Graham, E., Lee, M., Hill, R., Abbott, C., and Gordon, C.(1997). Expression and mapping of the mouse S7/*Pmsc2* gene, homolog of an essential mitotic gene in yeast. *Mamm. Genome* **8**, 352-354.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J.(1990). Basic local alignment search tool. *J. Mol. Biol.* **215**, 403-410.

Antequera, F. and Bird, A. P.(1993). Number of CpG islands and genes in human and mouse. *Proc. Nat. Acad. Sci.* **90**, 11995-11999.

Antoch, M. P., Song, E. -J., Chang, A. -M., Vitaterna, M. H., Zhao, Y., Wilsbacher, L. D., Sangoram, A. M., King, D. P., Pinto, L. H., and Takahashi, J. S.(1997). Functional identification of the mouse circadian *Clock* gene by transgenic BAC rescue. *Cell* **89**, 655-667.

Arveiler, B. and Porteous, D. J.(1991). Amplification of end fragments of YAC recombinants by inverse-polymerase chain reaction. *Technique* **3**, 24-28.

Avner, P., Amar, L., Dandolo, L., and Guenet, J. L.(1988). Genetic analysis of the mouse using interspecific crosses. *Trends Genet.* **4**, 18-23.

Aza-blanc, P., Ramirez-Weber, F. -A., Laget, M. -P., Schwartz, C., and Kornberg, T. B.(1997). Proteolysis that is inhibited by hedgehog targets Cubitus interruptus protein to the nucleus and converts it to a repressor. *Cell* **89**, 1043-1053.

Barsh, G. S. and Epstein, C. J.(1989). The long range restriction map surrounding the mouse agouti locus reveals a disparity between physical and genetic distances. *Genomics* **5**, 9-18.

Barton, D. E., Yang-Feng, T. L., Mason, A. J., Seeburg, P. H., and Franke, U.(1989). Mapping of genes for inhibin subunits a, $b_A$, and $b_B$ on human and mouse chromosomes and studies of *jsd* mice. *Genomics* **5**, 91-99.

Basler, K. and Struhl, G.(1994). Compartment boundaries and the control of *Drosophila* limb pattern by hedgehog protein. *Nature* **368**, 208-214.

Bellefroid, E. J., Lecocq, E. J., Benhida, A., Poncelet, D. A., Belayew, A., and Martial, J. A.(1989). The human genome contains hundereds of genes coding for finger proteins of the Kruppel type. *DNA* **8**, 377-387.

Belloni, E., Muenke, M., Roessler, E., Traverso, G., Siegelbartlet, J., Frumkin, A., Mitchel, H. F., Doniskeller, H., Helms, C., Hing, A. V., Heng, H. H. Q., Koop, B., Martindale, D., Rommens, J. M., Tsui, L. C., and Scherer, S. W.(1996). Identification of *Sonic hedgehog* as a candidate gene responsible for holoprosencephaly. *Nature Genet.* **14**, 353-356.

Bird, A. P.(1986). CpG-rich islands and the function of DNA methylation. *Nature* **321**, 209-213.

Bird, A. P.(1987). CpG islands as gene markers in the vertebrate nucleus. *Trends Genet.* **3**, 342-347.

Bird, A. P.(1989). Two classes of observed frequency for rare-cutter sites in CpG islands. *Nucleic Acids Res.* **17**, 9485.

Bird, A. P., Taggart, M., Frommer, M., Miller, O. J., and McLeod, D.(1985). A fraction of the mouse genome that is derived from islands of non-methylated CpG rich DNA. *Cell* **40**, 91-99.

Boehm, T.(1993). Analysis of multigene families by DNA fingerprinting of conserved domains: directed cloning of tissue-specific protein tyrosine phosphatases. *Oncogene* **8**, 1385-1390.

Bonhomme, F., Benmhedi, F., Britton-Davidian, J., and Martin, S.(1979). Analyse genetique de croisements interspecifique *Mus musculus L.* x *Mus spretus* Lataste: Liason de Adh1 avec Amy1 sur le chromosome 3 et de Es14 avec Mod1 sur le chromosome 9. *R. Acad. Sci. Paris* **289**, 545-548.

Brown, S. D. M. and Peters, J.(1996). Combining mutagenesis and genomics in the mouse - closing the phenotype gap. *Trends Genet.* **12**, 433-435.

Buchberg, A. M., Taylor, B. A., Jenkins, N. A., and Copeland, N. G.(1986). Chromosomal localisation of *Emv-16* and *Emv-17*, two closely linked ecotropic proviruses of RF/J mice. *J. Virol.* **60**, 1175-1178.

Buluwela, L., Forster, A., Boehm, T., and Rabbitts, T. H.(1989). A rapid procedure for colony screening using nylon filters. *Nucleic Acids Res.* **17**, 452.

Burke, A. C., Nelson, C. E., Morgan, B. A., and Tabin, C.(1995). *Hox* genes and the evolution of vertebrate axial morphology. *Development* **121**, 333-346.

Burke, D. T., Carle, G. F., and Olsen, M. V.(1987). Cloning of large segments of exogenous DNA into yeast by means of artificial chromosome vectors. *Science* **236**, 806-812.

Capecchi, M.(1989). The new mouse genetics: Altering the genome by gene targeting. *Trends Genet.* **5**, 70-76.

Carlson, B. M.(1988). *Patten's Foundations of Embryology* (5th edition. McGraw-Hill, 392-421.

Carter, T. C.(1954). The genetics of luxate mice. IV. Embryology. *J. Hered.* **52**, 1-35.

Carter, T. C.(1954). The genetics of luxate mice. IV. Embryology. *J. Genet.* **52**, 1-35.

Cattanach, B. M., Burtenshaw, M. D., Rasberry, C., and Evans, E. P.(1993). Large deletions and other gross forms of chromosome imbalance compatible with viability and fertility in the mouse. *Nature Genet.* **3**, 56-61.

Chan, D. C., Laufer, E., Tabin, C., and Leder, P.(1995). Polydactylous limbs in Strong's Luxoid mice result from ectopic polarizing activity. . *Development* **121**, 1971-1978.

Chang, D. T., Lopez, A., von Kessler, D. P., Chiang, C., Simandl, B. K., Zhao, R., Seldin, M. F., Fallon, J. F., and Beachy, P. A.(1994). Products, genetic linkage and limb patterning activity of a murine *hedgehog* gene. *Development* **120**, 3339-3353.

Charitie, J., De Graff, W., Shen, S., and Deschamps, J.(1994). Ectopic expression of *Hoxb8* causes duplication of the ZPA in the forelimb and homeotic transformation of axial structures. *Cell* **78**, 589-601.

Chartier, F. L., Keer, J. T., Sutcliffe, M. J., Henriques, D. A., Mileham, P., and Brown, S. D. M.(1992). Construction of a mouse yeast artificial chromosome library in a recombination-deficient strain of yeast. *Nature Genet.* **1**, 132-136.

Chen, P., Yu, Y. M., and Reddi, A. H.(1993). Chondrogenesis in chick limb bud mesodermal cells: Reciprocal modulation by inhibin and activin. *Exp. Cell Res.* **206**, 119-127.

Chiang, C., Litingtung, Y., Lee, E., Young, K. E., Corden, J. L., Westphal, H., and Beachy, P. A.(1996). Cyclopia and defective axial patterning in mice lacking *Sonic hedgehog* gene function. *Nature* **383**, 407-413.

Chou, Q., Russell, M., Birch, D. E., Raymond, J., and Bloch, W.(1992). Prevention of pre-PCR mis-priming and primer dimerisation improves low copy number amplifications. *Nucleic Acids Res.* **20**, 1717-1723.

Chowdhury, K., Deutsch, U., and Gruss, P.(1987). A multigene family encoding several "finger" structures is present and differentially active in mammalian genomes. *Cell* **48**, 771-778.

Ciechanover, A.(1994). The ubiquitin-proteasome proteolytic pathway. *Cell* **79**, 13-21.

Ciechanover, A., Heller, H., Elias, S., Haas, A. L., and Hershko, A.(1980). ATP-dependant conjugation of reticulocyte proteins with the polypeptide required for protein degradation. *Proc. Nat. Acad. Sci.* **77**, 1365-1368.

Clarren, S. K., Alvord, E. C., and Hall, J. G.(1980). Congenital hypothalamic hamartoblastoma, hypopituitarism, imperforate anus and postaxial polydactyly-A new syndrome? Part II: Neuropathological considerations. *Am. J. Med. Gen.* **7**, 75-83.

Cline, J., Braman, J. C., and Hogrefe, H. H.(1996). PCR fidelity of *Pfu* DNA polymerase and other thermostable DNA polymerases. *Nucleic Acids Res.* **24**, 3546-3551.

Cohn, M. J., Izpisua-Belmonte, J. -C., Abud, H., Heath, J. K., and Tickle, C.(1995). Fibroblast growth factors induce additional limb development from the flank of chick embryos. *Cell* **80**, 239-246.

Cohn, M. J., Patel, K., Krumlauf, R., Wilkinson, D. G., Clarke, J. G. W., and Tickle, C.(1997). *Hox9* genes and vertebrate limb specification. *Nature* **387**, 97-101.

Cohn, M. J. and Tickle, C.(1996). Limbs: a model for pattern formation within the vertebrate body plan. *Trends Genet.* **12**, 253-257.

Collins, F.(1995). Positional cloning moves from perditional to traditional. *Nature Genet.* **9**, 347-350.

Collins, F. S.(1992). Positional cloning: lets not call it reverse anymore. *Nature Genet.* **1**, 3-6.

Copeland, N. G., Jenkins, N. A., Gilbert, D. J., Eppig, J. T., Maltais, L. J., Miller, J. C., Dietrich, W. F., Weaver, A., Lincoln, S. E., Steen, R. G., Stein, L. D., Nadeau, J. H., and Lander, E,S.(1993). A genetic linkage map of the mouse: current applications and future prospects. *Science* **262**, 57-66.

Crossley, P. H. and Martin, G. R.(1995). The mouse *Fgf8* gene encodes a family of polypeptides and is expressed in regions that direct outgrowth and patterning in the developing embryo. *Development* **121**, 439-451.

Crossley, P. H., Minowada, G., MacArthur, C. A., and Martin, G. R.(1996). Roles for FGF8 in the induction, initiation, and maintenance of chick limb development. *Cell* **84**, 127-136.

Danielson, P. E., Watson, J. B., Gerendasy, D. D., Erlander, M. G., Lovenberg, T. W., Delecea, L., Sutcliffe, J. G., and Frankel, W. N.(1994). Chromosomal mapping of mouse genes expressed selectively within the central nervous system. *Genomics* **19**, 454-461.

Darling, S. M. and Abbott, C. M.(1992). Mouse models of human gene disorders I: Non-transgenic mice. *Bioessays* **14**, 359-366.

Davidson, D., Graham, E., Sime, C., and Hill, R.(1988). A gene with sequence similarity to *Drosophila engrailed* is expressed during the development of the neural tube and vertebrae in the mouse. *Development* **104**, 305-316.

Davidson, D. R., Crawley, A., Hill, R. E., and Tickle, C.(1991). Position-dependent expression of two related homeobox genes in developing vertebrate limbs. *Nature* **352**, 429-431.

Davis, A. P. and Capecchi, M. R.(1994). Axial homeosis and appendicular skeleton defects in mice with a targeted disruption of *Hoxd11*. *Development* **120**, 2187-2198.

Davis, A. P., Witte, D. P., Hsieh-li, H. M., Potter, S. S., and Capecchi, M. R.(1995). Absence of radius and ulna in mice lacking *Hoxa11* and *Hoxd11*. *Nature* **375**, 791-795.

Davis, C. A., Holmyard, D. P., Millen, K. J., and Joyner, A. L.(1991). Examining pattern formation in mouse, chicken and frog embryos with an *En* specific antisera. *Development* **111**, 287-298.

Davis, C. A. and Joyner, A. L.(1988). Expression patterns of the homeobox containing genes *En-1* and *En-2* and the proto-oncogene *int-1* diverge during development. *Genes & Dev.* **2**, 1736-1744.

Dawson, S. P., Arnold, J. E., Mayer, N. J., Reynolds, S. E., Billett, M. A., Gordon, C., Colleaux, L., Kloetzel, P. M., Tanaka, K., and Mayer, R. J.(1995). Developmental changes of the 26S proteasome in abdominal intersegmental muscles of *Manduca sexta* during programmed cell death. *J. Biol. Chem.* **270**, 1850-1858.

Dealy, C. N., Roth, A., Ferrari, D., Brown, A. M. C., and Kosher, R. A.(1993). *Wnt5a* and *Wnt7a* are expressed in the developing chick limb bud in a manner suggesting roles in pattern formation along the proximodistal and dorsoventral axes. *Mech. Dev.* **43**, 175-186.

Dear, T. N., Colledge, W. H., Carlton, M. B. L., Lavenir, I., Larson, T., Smith, A. J. H., Warren, A. J., Evans, M. J., Sofroniew, M. V., and Rabbitts, T. H.(1995). The *Hox11* gene is essential for cell survival during spleen development. *Development* **121**, 2909-2915.

Deveraux, Q., Ustrell, V., Pickart, C., and Rechsteiner, M.(1994). A 26S protease subunit that binds ubiquitin conjugates. *J. Biol. Chem.* **269**, 7059-7061.

Devon, R. S., Porteous, D. J., and Brookes, A. J.(1995). Splinkerettes, improved vectorettes for greater efficiency in PCR walking. *Nucleic Acids Res.* **23**, 1644-1645.

Dice, J. F. and Terlecky, S. R.(1993). Polypeptide import and degradation by isolated lysosomes. *J. Biol. Chem.* **268**, 23490-23495.

Dickie, M. M.(1968). . *Mouse News Lett.* **38**, 24.

Dietrich, W., Katz, H., Lincoln, S. E., Shin, H. -S., Friedman, J., Dracopoli, N. C., and Lander, E. S.(1992). A genetic map of the mouse suitable for typing intraspecific crosses. *Genetics* **131**, 423-447.

Dietrich, W. F., Copeland, N. G., Gilbert, D. J., Miller, J. C., Jenkins, N. A., and Lander, E. S.(1995). Mapping the mouse genome: Current status and future prospects. *Proc. Nat. Acad. Sci.* **92**, 10849-10853.

Dietrich, W. F., Miller, J. C., Steen, R. G., Merchant, M., Damron, D., Nahf, R., Gross, A., Joyce, D. C., Wessel, M., Dredge, R. D., Marquis, A., Stein, L. D., Goodman, N., Page, D. C., and Lander, E. S.(1994). A genetic map of the mouse with 4,006 simple sequence length polymorphisms. *Nature Genet.* **7**, 220-245.

Dolle, P., Dierich, A., Lemeur, M., Schimmang, T., Schuhbaur, B., Chambon, P., and Duboule, D.(1993). Disruption of the *Hoxd13* gene induces localised heterochrony leading to mice with neotenic limbs. *Cell* **75**, 431-441.

Dominguez, M., Brunner, M., Hafen, E., and Basler, K.(1996). Sending and recieving the Hedgehog signal: Control by the *Drosophila* Gli protein Cubitus interruptus. *Science* **272**, 1621-1625.

Dubiel, W., Ferrel, K., and Rechsteiner, M.(1993). Peptide sequencing identifies a modulator of HIV Tat-mediated transactivation, as subunit 7 of the 26S protease. *FEBS Lett.* **323**, 276-278.

Dubiel, W., Ferrell, K., Pratt, G., and Rechsteiner, M.(1992). Subunit-4 of the 26S protease is a member of a novel eukaryotic ATPase family. *J. Biol. Chem.* **267**, 22699-22702.

E.U.C.I.B.Group, (1994). Towards high resolution maps of the mouse and human genomes - a facility for ordering markers to 0.1 cM resolution. *Hum. Mol. Genet.* **3**, 621-627.

Elgar, G., Sandford, R., Aparicio, S., MacRae, A., Venkatesh, B., and Brenner, S.(1996). Small is beautiful: comparative genomics with the pufferfish (Fugu rubripes). *Trends Genet.* **12**, 145-150.

Epstein, D. J., Vekemans, M., and Gros, P.(1991). *Splotch* ($Sp^{2h}$), a mutation affecting development of the mouse neural tube, shows a deletion within the paired homeodomain of *Pax3*. *Cell* **67**, 767-774.

Erickson, R. P.(1989). Why isn't a mouse more like a man. *Trends Genet.* **5**, 1-3.

Esch, F. S., Shimasaki, S., Cooksey, K., Mercado, M., Mason, A. J., Ying, S. -Y., Ueno, N., and Ling, N.(1987). Complementary deoxyribonucleic acid (cDNA) cloning and DNA sequence analysis of rat ovarian inhibins. *Mol. Endocrinol.* **1**, 388-396.

Fahrbach, S. E. and Schwartz, L. M.(1994). Localisation of immunoreactive ubiquitin in the nervous system of the *Manduca sexta* moth. *J. Comp. Neurol.* **343**, 464-482.

Fallon, J. F. and Kelley, R. O.(1977). Ultrastructural analysis of the apical ectodermal ridge during vertebrate limb morphogenesis. II Gap junctions as distinctive ridge structures common to birds and mammals. *J. Embryol. Exp. Morph.* **41**, 223-232.

Fan, C. -M., Porter, J. A., Chiang, C., Chang, D. T., Beachy, P., and Tessier-Lavigne, M.(1995). Long range sclerotome induction by *Sonic hedgehog*: direct role of the amino terminal cleavage product and modulation by the cyclic AMP signalling pathway. *Cell* **81**, 457-465.

Fang, J., Wang, S. -Q., Smiley, E., and Bonadio, J.(1997). Genes coding for mouse activin $b_C$ and $b_E$ are closely linked and exhibit a liver-specific expression pattern in adult tissues. *Biochem. Biophys. Res. Comm.* **231**, 655-661.

Fang, J. M., Yin, W. S., Smiley, E., Wang, S. Q., and Bonadio, J.(1996). Molecular cloning of the mouse activin b(E) subunit gene. *Biochem. Biophys. Res. Comm.* **228**, 669-674.

Fantes, J., Oghene, K., Boyle, S., Danes, S., Fletcher, J., Bruford, E., Williamson, K., Seawright, A., Schedl, A., Hanson, I., Zehtner, G., Bhogal, R., Lehrach, H., Gregory, S., Williams, S., Little, P., Sellar, G., Hoovers, J., Mannens, M., Weissbach, J., Jumein, C., van Heyningen, V., and Bickmore, W.(1995). A high resolution integrated physical, cytogenetic and genetic map of human chromosome 11: distal p13 to proximal p15.1. *Genomics* **25**, 447-461.

Favier, B., Lemeur, M., Chambon, P., and Dolle, P.(1995). Axial skeleton homeosis and forelimb malformations in *Hoxd11* mutant mice. *Proc. Nat. Acad. Sci.* **92**, 310-314.

Feijen, A., Goumans, M. J., and van den Eijnden-van Raaij, A. J. M.(1994). Expression of activin subunits, activin receptors and follistatin in postimplantation mouse embryos suggests specific developmental functions for different activins. *Development* **120**, 3621-3637.

Feng, Z. -M., Li, Y. -P., and Chen, C. -L. C.(1989). Analysis of the 5'-flanking regions of rat inhibin a- and b-B-subunit genes suggests two different regulatory mechanisms. *Mol. Endocrinol.* **3**, 1914-1925.

Forsthoefel, P.(1959). The embryological development of the skeletal effects of the luxoid gene in the mouse, including its interactions with the luxate gene. *J. Morphol.* **104**, 89-141.

Francis, P. H., Richardson, M. K., Brickell, P. M., and Tickle, C.(1994). Bone morphogenetic proteins and a signalling pathway that controls patterning in the developing chick limb. *Development* **120**, 209-218.

Francis-West, P. H., Robertson, K. E., Ede, D. A., Rodriguez, C., Izpisua-Belmonte, J. C., Houston, B., Burt, D. W., Gribbin, C., Brickell, P. M., and Tickle, C.(1995). Expression of genes encoding bone morphogenic proteins and sonic hedgehog in *talpid* ($ta^3$) limb buds - Their relationships in the signalling cascade involved in limb patterning. *Dev. Dyn.* **203**, 187-197.

Frankel, W. N., Lee, B. K., Stoye, J. P., Coffin, J. M., and Eicher, E. M.(1992). Characterisation of the endogenous nonecotropic murine leukemia viruses of NZB/B1NJ and SM/J inbred strains. *Mamm. Genome* **2**, 110-122.

Frankel, W. N., Stoye, J. P., Taylor, B. A., and Coffin, J. M.(1989). Genetic analysis of endogenous xenotropic murine leukemia viruses: Association with 2 common mouse mutations and the viral restriction locus *Fv-1*. *J. Virol.* **63**, 1763-1774.

Franz, R. and Besecke, A.(1991). The development of the eye in homozygotes of the mouse mutant Extra toes. *Anat Embryol.* **184**, 355-361.

Franz, T.(1994). *Extra toes* (*Xt*) homozygous mutant mice demonstrate a role for the *Gli3* gene in the development of the forebrain. *Acta Anat.* **150**, 38-44.

Friedrich, G. and Soriano, P.(1991). Promotor traps in embryonic stem cells: A genetic screen to identify and mutate developmental genes in mice. *Genes & Dev.* **5**, 1513-1523.

Frohman, M. A., Dickinson, M. E., Hogan, B. L. M., and Martin, G. R.(1993). Mapping of *Gbx-1* to mouse chromosome 5 and *Gbx-2* to mouse chromosome1. *Mouse genome* **91**, 323-325.

Geduspan, J. S. and Solursh, M.(1992). A growth promoting influence from the mesonephros during limb outgrowth. *Dev. Biol.* **151**, 212-250.

Genetics Computer Group, (1994). *Program manual for the Wisconsin package*, 8th edition. 575, Science Drive Madison, Wisconsin, USA, 53711.

Ghislain, M., Udvardy, A., and Mann, C.(1993). *S. cerevisiae* 26S protease mutants arrest cell division in G2/metaphase. *Nature* **366**, 358-362.

Gibson, F., Walsh, J., Mburu, P., Varela, A., Brown, K. A., Antonio, M., Beisel, K. W., Steel, K. P., and Brown, S. D. M.(1995). A type-VII myosin encoded by the mouse deafness gene *shaker-1*. *Nature* **374**, 62-64.

God *et al.*, *The Holy Bible*, (1817). Clarendon Press, Oxford.

Goodrich, L. V., Johnson, R. L., Milenkovic, L., McMahon, J. L., and Scott, M. P.(1996). Conservation of the *hedgehog/patched* signalling pathway from flies to mice: induction of a mouse *patched* gene by Hedgehog. *Genes & Dev.* **10**, 301-312.

Gordon, C., McGurk, G., Dillon, P., Rosen, C., and Hastie, N. D.(1993). Defective mitosis due to a mutation in the gene for a fission yeast 26S protease subunit. *Nature* **366**, 355-357.

Gossler, A., Joyner, A. L., Rossant, J., and Skarnes, W. C.(1989). Mouse embryonic stem cells and reporter constructs to detect developmentally regulated genes. *Science* **244**, 463-465.

Green, E. D., Riethman, H. C., Dutchik, J. E., and Olsen, M. V.(1991). Detection and characterisation of chimeric yeast artificial chromosome clones. *Genomics* **11**, 658-669.

Green, E. L. and Roderick, T. H.(1966). *Biology of the laboratory mouse* (Green, E. L., Ed.) McGraw-Hill, New York. 165-185.

Green, M. C.(1967). A defect of the splanchnic mesoderm caused by the mutant gene *Dominant hemimelia* in the mouse. *Dev. Biol.* **15**, 62-89.

Green, M. C.(1989). *Genetic Variants and Strains of the Laboratory Mouse* (Lyon, M. F. and Searle, A. G., Eds.) 2nd edition. Oxford University Press, Oxford. 12-403.

Greig, D. M.(1926). Oxycephaly. *Edinburgh Med. J.* **33**, 189-218.

Grunstein, M. and Hogness, D. S.(1997). Colony hybridisation: A method for the isolation of cloned DNAs that contain a specific gene. *Proc. Nat. Acad. Sci.* **72**, 3961-3964.

Gu, H., Marth, J. D., Orban, P. C., Mossmann, H., and Rajewsky, K.(1994). Deletion of a DNA polymerase-beta gene segment in T-cells using cell-type specific gene targeting. *Science* **265**, 103-106.

Hahn, H., Wicking, C., Zaphiropoulos, P. G., Gailani, M. R., Shanley, S., Chidambaram, A., Vorechovsky, I., Holmberg, E., Unden, A. B., Gillies, S., Negus, K., Smyth, I., Pressman, C., Leffell, D. J., Gerrard, B., Goldstein, A. M., Dean, M., Toftgard, R., Chenevixtrench, G., Wainwright, B., and Bale, A. E.(1996). Mutations of the human homologue of *Drosophila* patched in the nevoid basel cell carcinoma syndrome. *Cell* **85**, 841-851.

Hall, J. G., Pallister, P. D., Clarren, S. K., Beckwith, J. B., Wiglesworth, F. W., Fraser, F. C., Cho, S., Benke, P. J., and Reed, S. D.(1980). Congenital hypothalamic hamartoblastoma, hypopituitarism, imperforate anus and postaxial polydactyly-A new syndrome? Part I: Clinical, causal and pathogenetic considerations. *Am. J. Med. Gen.* **7**, 47-74.

Hammer, M. F., Schimenti, J., and Silver, L. M.(1989). Evolution of mouse chromosome 17 and the origin of inversions associated with *t* haplotypes. *Proc. Nat. Acad. Sci.* **86**, 3261-3265.

Hammerschmidt, M., Bitgood, M. J., and McMahon, A. P.(1996). Protein kinase A is a common negative regulator of Hedgehog signalling in the vertebrate embryo. *Genes & Dev.* **10**, 647-658.

Hammerschmidt, M., Brook, A., and McMahon, A. P.(1997). The world according to *hedgehog*. *Trends Genet.* **13**, 14-21.

Hamvas, R. M. J, Francis, F., Cox, R. D., Nizetic, D., Goldsworthy, M. E., Brown, S. D. M., and Lehrach, H. R.(1994). Rapid restriction analysis of YAC clones. *Nucleic Acids Res.* **22**, 1318-1319.

Hashimoto, M., Shoda, A., Inoue, S., Yamada, R., Kondo, T., Sakurai, T., Ueno, N., and Murmatsu, M.(1992). Functional regulation of osteoblastic cells by the interaction of activin A with follistatin. *J. Biol. Chem.* **267**, 4999-5004.

Hayamizu, T. F., Sessions, S. K., Wanek, N., and Bryant, S. V.(1991). Effects of localised application of transforming growth factor bI on developing chick limbs. *Dev. Biol.* **145**, 164-173.

Hayashizaki, Y., Hirotsune, S., Okazaki, Y., Shibata, H., Akasako, A., Muramatsu, M., Kawai, J., Hirasawa, T., Watanabe, S., Shiroishi, T., Moriwaki, K., Taylor, B. A., Matsuda, Y., Elliott, R. W., Manly, K. F., and Chapman, V. M.(1994). A genetic linkage map of the mouse using restriction landmark genomic scanning (RLGS). *Genetics* **138**, 1207-1238.

Heikinheimo, M., Lawshe, A., Shackleford, G. M., Wilson, D. B., and MacArthur, C. A.(1994). *Fgf8* expression in the post-gastrulation mouse suggests roles in the development of the face, limbs and central nervous system. *Mech. Dev.* **48**, 129-138.

Heiskanen, M., Peltonen, L., and Palotie, A.(1996). Visual mapping by high resolution FISH. *Trends Genet.* **12**, 379-382.

Hendrick, J. P. and Hartl, F-U.(1993). Molecular chaperone functions of heat shock proteins. *Ann. Rev. Biochem.* **62**, 349-384.

Hendrick, P. W. and Levin, D. A.(1984). Kin-founding and the fixation of chromosomal variants. *Am. Nat.* **124**, 789-797.

Herman, G. E., Nadeau, J. H., and Hardies, S. C.(1992). Dispersed repetitive elements in mouse genome analysis. *Mamm. Genome* **2**, 207-214.

Hershko, A., Ciechanover, A., Heller, H., Haas, A. L., and Rose, I. A.(1980). Proposed role of ATP in protein breakdown: Conjugation of protein with multiple chains of the polypeptide of ATP-dependant proteolysis. *Proc. Nat. Acad. Sci.* **77**, 1783-1786.

Heutink, P., Zguricas, J., Oosterhout, L. v., Breedveld, G. J., Testers, L., Sandkuijl, L. A., Snijders, P. J. L. M., Weissenbach, J., Lindhout, D., Hovius, S. E. R., and Oostra, B. A.(1994). The gene for triphalangeal thumb maps to the subtelomeric region of chromosome 7q. *Nature Genet.* **6**, 287-291.

Higgins, M.(1991). *Genetic and molecular studies of the Dominant hemimelia locus in the mouse. Ph.D. Thesis*, University of Edinburgh,

Higgins, M., Hill, R. E., and West, J. D.(1992). Dominant hemimelia and En-1 on mouse chromosome 1 are not allelic. *Genet. Res.* **60**, 53-60.

Hill, R. E., Hall, A. E., Sime, C. M., and Hastie, N. D.(1987). A mouse homeo box-containing gene maps near a developmental mutation. *Cytogenet. Cell Genet.* **44**, 171-174.

Hill, R. E., Jones, P. F., Rees, A. R., Sime, C. R., Justice, M. J., Copeland, N. G., Jenkins, N. A., Graham, E., and Davidson, D. R.(1989). A new family of mouse homeo box-containing genes: molecular structure, chromosomal location, and developmental expression of Hox-7.1. *Genes & Dev.* **3**, 26-37.

Hillyard, A. L., Doolittle, D. P., Davisson, M. T., and Roderick, T. H.(1992). Locus map of mouse. *Mouse genome* **90**, 8-21.

Hinchliffe, J. R. and Sansom, A.(1985). The distribution of the polarizing zone (ZPA) in the leg-bud of the chick embryo. *J. Embryol. Exp. Morph.* **86**, 169-175.

Hing, A. V., Helms, C., Slaugh, R., Burgess, A., Wang, J. C., Herman, T., Dowton, S. B., and Doniskeller, H.(1995). Linkage of preaxial polydactyly type-2 to 7q36. *Am. J. Med. Gen.* **58**, 128-135.

Honig, L. S. and Summerbell, D.(1985). Maps of position strength of positional signalling activity in the developing chick wing. *J. Embryol. Exp. Morph.* **87**, 163-174.

Hooper, J. and Scott, M. P.(1989). The *Drosophila patched* gene encodes a putative membrane protein required for segmental patterning. *Cell* **59**, 751-765.

Hornbruch, A. and Wolpert, L.(1991). The spatial and temporal distribution of polarizing activity in the flank of pre-limb-bud stages in the chick embryo. *Development* **111**, 725-731.

Hotten, G., Neidhardt, H., Schneider, C., and Pohl, J.(1995). Cloning of a new member of the TGF-b family: A putitive new activin$b_c$ chain. *Biochem. Biophys. Res. Comm.* **206**, 608-613.

Hough, R., Pratt, G., and Rechsteiner, M.(1987). Purification of two high molecular weight proteases from rabbit reticulocyte lysate. *J. Biol. Chem.* **262**, 8303-8313.

Hughes, D. C., Allen, J., Morley, G., Sutherland, K., Ahmed, W., Prosser, J., Lettice, L., Allan, G., Mattei, M-G., Farrall, M., and Hill, R. E.(1997). Cloning and sequencing of the mouse Gli2 gene: Localisation to the Dominant hemimelia critical region. *Genomics* **39**, 205-215.

Hui, C-C., Slusarski, D., Platt, K. A., Holmgren, R., and Joyner, A. L.(1994). Expression of three Mouse homologues of the *Drosophila* segment polarity gene *cubitus interuptus*, *Gil*, *Gli-2* and *Gli-3*, in ectoderm and mesoderm derived tissues suggests multiple roles during postimplantation development. *Dev. Biol.* **162**, 402-413.

Hui, C. -c. and Joyner, A. L.(1993). A mouse model of Greig cephalopolysyndactyly syndrome: the extra-toes$^J$ mutation contains an intragenic deletion of the Gli3 gene. *Nature Genet.* **3**, 241-245.

Hynes, M., Porter, J. M., Chiang, C., Chang, D., Tessier-Lavigne, M., Beachy, P. A., and Rosenthal, A.(1995). Induction of midbrain dopaminergic neurons by Sonic hedgehog. *Neuron* **15**, 35-44.

Irvine, K. D. and Wieschaus, E.(1994). *fringe*, a boundary-specific signalling molecule, mediates interactions between dorsal and ventral cells during *Drosophila* wing development. *Cell* **79**, 595-606.

Ispisua-Belmonte, J. -C. and Duboule, D.(1992). Homeobox genes and pattern formation in the vertebrate limb. *Dev. Biol.* **152**, 26-36.

Itoh, M., Igarashi, M., Yamada, K., Hasegawa, Y., Seki, M., Eto, Y., and Shibai, H.(1990). Activin A stimulates meiotic maturation of the rat oocyte in vitro. *Biochem. Biophys. Res. Comm.* **166**, 1479-1484.

Jaenisch, R.(1988). Transgenic animals. *Science* **240**, 1468-1474.

Jenkins, N. A., Copeland, N. G., Taylor, B. A., Bedigian, H. G., and Lee, B. K.(1982). Ecotropic murine leukemia virus DNA content of normal and lymphomatous tissues of BXH-2 recombinant inbred mice. *J. Virol.* **42**, 379-388.

Jentsch, S.(1992). The ubiquitin-conjugation system. *Annu. Rev. Genet.* **26**, 179-207.

Jiang, T. X., Yi, J. R., Ying, S. Y., and Chuong, C. M.(1993). Activin enhances chondrogenesis of limb bud cells: Stimulation of precartilaginous mesenchymal condensations and expression of NCAM. *Dev. Biol.* **155**, 545-557.

Johnson, D. R.(1967). *Extra toes*: a new mutant gene causing multiple abnormalities in the mouse. *J. Embryol. Exp. Morph.* **17**, 543-581.

Johnson, R. L., Grenier, J., and Scott, M. P.(1995). Patched overexpression alters wing disc size and pattern: transcriptional and post-transcriptional effects on hedgehog targets. *Development* **121**, 4161-4170.

Johnson, R. L., Rothman, A. L., Xie, J., Goodrich, L. V., Bare, J. W., Bonifas, J. M., Quinn, A. G., Myers, R. M., Cox, D. R., Epstein Jr., E. H., and Scott, M. P.(1996). Human homologue of *patched*, a candidate gene for the basal cell nevus syndrome. *Science* **272**, 1668-1671.

Joyner, A. L., Auerbach, A., and Skanes, W. C.(1992). The gene trap approach in embryonic stem cells: the potential for genetic screens in mice. *Ciba Found. Symp.* **165**, 277-288.

Joyner, A. L. and Martin, G. R.(1987). *En-1* and *En-2*, two mouse genes with sequence homology to the *Drosophila engrailed* gene: expression during embryogenesis. *Genes & Dev.* **1**, 29-38.

Joyner, A.L., ed. *Gene Targeting: A Practical Approach*, (1993). Oxford University Press, New York.

Justice, M. J., Siracusa, L. D., Gilbert, D. J., Heisterkamp, N., Groffen, J., Chada, K., Silan, C. M., Copeland, M. G., and Jenkins, N. A.(1990). A genetic linkage map of mouse chromosome 10: Localisation of eighteen molecular markers using a single interspecific backcross. *Genetics* **125**, 855-866.

Kang, S., Graham Jr., J. M., Olney, A. H., and Biesecker, L. G.(1997). *GLI3* frameshift mutations cause autosomal dominant Pallister-Hall syndrome. *Nature Genet.* **15**, 266-268.

Katsanis, N., Fitzgibbon, J., and Fisher, E. M. C.(1996). Paralogy mapping: Identification of a region in the human MHC triplicated onto human chromosomes 1 and 9 allows the prediction and isolation of novel *PBX* and *NOTCH* loci. *Genomics* **35**, 101-108.

Kieny, M.(1968). Variation de la capacite inductrice du mesoderme et de la competence de l'ectoderme au cours de l'induction primaire du bourgeon de membre chez l'embryon de poulet. *Morphol. Exp.* **57**, 401-418.

Kim, J., Irvine, K. D., and Carroll, S. B.(1995). Cell recognition, signal induction, and symmetrical gene activation at the dorso-ventral boundary of the developing *Drosophila* wing. *Cell* **82**, 795-802.

Kim, Y-J., Bjorkland, S., Li, Y., Sayre, M. H., and Kornberg, R. D.(1994). A multiprotein mediator of transcriptional activation and its interaction with the C-terminal repeat domain of RNA polymerase II. *Cell* **77**, 599-608.

Kinzler, K. W., Bigner, S. H., Bigner, D. D., Trent, J. M., Law, M. L., O'Brien, S. J., Wong, A. J., and Vogelstein, B.(1987). Identification of an amplified, highly expressed gene in a human glioma. *Science* **236**, 70-73.

Kinzler, K. W., Ruppert, J. M., Bigner, S. H., and Vogelstein, B.(1988). The Gli gene is a member of the Kruppel family of zinc finger proteins. *Nature* **332**, 371-374.

Kinzler, K. W. and Vogelstein, B.(1990). The GLI gene encodes a nuclear protein which binds specific sequences in the human genome. *Mol. Cell Biol.* **10**, 634-642.

Knoth, K., Roberds, S., Poteet, C., and Tamkun, M.(1988). Highly degenerate, inosine containing primers specifically amplify rare cDNA using the polymerase chain reaction. *Nucleic Acids Res.* **16**, 10932.

Knudsen, T. B. and Kochhar, D. M.(1981). The role of morphogenetic cell death during abnormal limb-bud outgrowth in mice heterozygous for the dominant mutation *Hemimelia-extra toe* (*Hm*$^{X}$). *Embryol. Exp. Morphol.* **65**, 289-307.

Kozak, C. A., Bucan, M., Goffinet, A., and Stephenson, D. A.(1996). Mouse chromosome 5. *Mamm. Genome* **6**, s97-s112.

Kruger, G., Gotz, J., Kvist, U., Dunker, H., Erfurth, F., Pelz, L., and Zech, L.(1989). Greig syndrome in a large kindred due to reciprocal chromosome translocation t(6;7)(q27;p13). *Am. J. Med. Gen.* **32**, 411-416.

Kuhn, R., Schwenk, F., Aguet, M., and Rajewsky, K.(1995). Inducible gene targeting in mice. *Science* **269**, 1427-1429.

Kusumi, K., Smith, J., Segre, J., Koos, D., and Lander, E.(1993). Construction of a large insert yeast artificial chromosome library of the mouse genome. *Mamm. Genome* **4**, 391-392.

Lafuse, W. P. and Zwilling, B. S.(1993). Localization of the inhibin $b_B$ gene on mouse chromosome 1. *Mamm. Genome* **4**, 399-400.

Larin, Z., Monaco, A. P., and Lehrach, H.(1991). Yeast artificial chromosome libraries containing large inserts from mouse and human DNA. *Proc. Nat. Acad. Sci.* **88**, 4123-4127.

Larsen, F., Gunderson, G., Lopez, R., and Prydz, H.(1992). CpG islands as gene markers in the human genome. *Genomics* **13**, 1095-1107.

Lau, A. L., Nishimori, K., and Matzuk, M. M.(1996). Structural analysis of the mouse activin $b_C$ gene. *Biochem. Biophys. Acta* **1307**, 145-148.

Laufer, E., Dahn, R., Orozco, O. E., Yeo, C. Y., Pisenti, J., Henrique, D., Abbott, U. K., Fallon, J. F., and Tabin, C.(1997). Expression of *Radical fringe* in limb bud ectoderm regulates apical ectodermal ridge formation. *Nature* **386**, 366-373.

Laufer, E., Nelson, C. E., Johnson, R. I., Morgan, B. A., and Tabin, C.(1994). *Sonic hedgehog* and *Fgf4* act through a signalling cascade and feedback loop to integrate growth and patterning of the developing limb bud. *Cell* **79**, 993-1003.

Lee, C. C., Wu, X., Gibbs, R. A., Cook, R. G., Muzny, D. M., and Caskey, C. T.(1988). Generation of cDNA probes directed by amino acid sequence: Cloning of urate oxidase. *Science* **239**, 1288-1291.

Li, W. -H.(1982). *Isozymes: Current topics in biological and medical research* (Rattazzi, M. C., Scandalios, J. G., and Whitt, G. S., Eds.) 6th edition. A. R. Liss, New York. 55-92.

Lin, H. Y., Wang, X. -F., Ng-Eaton, E., Weinberg, R. A., and Lodish, H. F.(1992). Expression cloning of the TGF-b type II receptor, a functional transmembrane serine/threonine kinase. *Cell* **68**, 775-785.

Lindsay, S. and Bird, A. P.(1987). Use of restriction enzymes to detect potential gene sequences in mammalian DNA. *Nature* **327**, 336-338.

Lock, L. F., Gilbert, D. J., Street, V. A., Migeon, M. B., Jenkins, N. A., Copeland, N. G., and Tempel, B. L.(1994). Voltage-gated potassium channel genes are clustered in paralogous regions of the mouse genome. *Genomics* **20**, 354-362.

Logan, C., Hanks, M. C., Noble-Topham, S., Nallainathan, D., Provart, N. J., and Joyner, A. L.(1992). Cloning and sequence comparrison of the mouse, human and chicken *engrailed* genes reveal potential functional domains and regulatory regions. *Dev. Genet.* **13**, 345-358.

Logan, C., Hornbruch, A., Campbell, I., and Lumsden, A.(1997). The role of *Engrailed* in establishing the dorsoventral axis in the chick limb. *Development* **124**, 2317-2324.

Loomis, C. A., Harris, C., Michaud, J., Wurst, W., Hanks, M., and Joyner, A. L.(1996). The mouse *Engrailed-1* gene and ventral limb patterning. *Nature* **382**, 360-363.

Love, J. M., Knight, A. M., McAleer, M. A., and Todd, J. A.(1990). Towards construction of a high resolution map of the mouse genome using PCR-analysed microsatellites. *Nucleic Acids Res.* **18**, 4123-4130.

Lowe, J., Stock, D., Jap, B., Zwickl, P., Baumeister, W., and Huber, R.(1995). Crystal structure of the 20S proteasome from the archaeon *T. acidophilum* at 3.4 A resolution. *Science* **268**, 533-539.

Lundin, L,-G.(1979). Evolutionary conservation of large chromosomal segments reflected in mammalian gene maps. *Clin. Genet.* **16**, 72-81.

Lundin, L,-G.(1989). Gene homologies with emphasis on paralogous genes and chromosomal regions. *Life Sci. Adv. (Genet. )* **8**, 89-104.

Lundin, L,-G.(1993). Evolution of the vertebrate genome as reflected in paralogous chromosomal regions in man and the house mouse. *Genomics* **16**, 1-19.

MacCabe, J. A., Errick, J., and Saunders, J. W. Jr.(1974). Ectodermal control of the dorsoventral axis in the leg bud of the chick embryo. *Dev. Biol.* **39**, 69-82.

Manly, K. F.(1993). A Mackintosh programme for storage and analysis of experimental genetic mapping data. *Mamm. Genome* **4**, 303-313.

Margio, V., Davey, R. A., Zou, Y., Cunningham, J. M., and Tabin, C. J.(1996). Biochemical evidence that patched is the Hedgehog receptor. *Nature* **384**, 176-179.

Marigo, V., Johnson, R. L., Vortkamp, A., and Tabin, C. J.(1996). Sonic hedgehog differentially regulates expression of *GLI* and *GLI3* during limb development. *Dev. Biol.* **180**, 273-283.

Martin, G. R., Richman, M., Reinsch, S., Nadeau, J. H., and Joyner, A. L.(1990). Mapping of the two mouse *engrailed*-like genes: close linkage of *En-1* to *dominant hemimelia*(*Dh*) on chromosome 1 and of *En-2* to *hemimelic extra toes* (*Hx*) on chromosome 5. *Genomics* **6**, 302-308.

Mason, A. J., Berkemeier, L. M., Schmelzer, C. H., and Schwall, R. H.(1989). Activin B: Precursor sequences, genomic structure and *in vitro* activities. *Mol. Endocrinol.* **3**, 1352-1358.

Masuya, H., Sagai, T., Moriwaki, K., and Shiroishi, T.(1997). Multigenic control of the localisation of the zone of polarising activity in limb morphogenesis in the mouse. *Dev. Biol.* **182**, 42-51.

Masuya, H., Sagai, T., Wakana, S., Moriwaki, K., and Shiroishi, T.(1995). A duplicated zone of polarising activity in polydactylous mouse mutants. *Genes & Dev.* **9**, 1645-1653.

Mathews, L. S.(1994). Activin receptors and cellular signaling by the receptor serine kinase family. *Endocr. Rev.* **15**, 310-325.

Matsui, T., Hirai, M., Hirano, M., and Kurosawa, Y.(1993). The *HOX* complex neighboured by the *EVX* gene, as well as two other homeobox-containing genes, the *GBX*-class and the *EN*-class, are located on the same chromosomes 2 and 7 in humans. *FEBS Lett.* **336**, 107-110.

Matthes, H., Boschert, U., Amlaiky, N., Grailhe, R., Plassat, J. L., Muscatelli, F., Mattei, M. -G., and Hen, R.(1993). Mouse 5-hydroxytriptamine$_{5A}$ and 5-hydroxytriptamine$_{5B}$ receptors define a new family of serotonin receptors: Cloning, functional expression and chromosomal location. *Mol. Pharm.* **43**, 313-319.

Matzuk, M. M., Kumar, T. R., Vassalli, A., Bickenbach, J. R., Roop, D. R., Jaenisch, R., and Bradley, A.(1995). Functional analysis of activins during mammalian development. *Nature* **374**, 354-356.

Matzuk, M. M., Lu, N., Vogel, H., Sellheyer, K., Roop, D. R., and Bradley, A.(1995). Multiple defects and perinatal death in mice deficient in follistatin. *Nature* **374**, 360-363.

McCusick, V. A.(1992). *Mendelian Inheritance in Man: Catalogues of Autosomal Dominant, Autosomal Recessive, and X-linked Phenotypes*, 10th edition. The John Hopkins University Press, Baltimore.

McKusick, V. A.(1992). *Mendelian Inheritance in Man*, 10th edition. John Hopkins University Press, Baltimore.

McPherson, M. J., Jones, K. M., and Gurr, S-J.(1991). *PCR. A practical approach, volume 1* (McPherson, M. J., Quirke, K., and Taylor, G. R., Eds.) IRL Press, Oxford. 171-186.

Meissler, M. H.(1992). Insertional mutation of classical and novel genes in transgenic mice. *Trends Genet.* **10**, 341-344.

Mian, I. S.(1993). Sequence similarities between cell regulation factors, heat shock proteins and RNA helicases. *Trends Biochem. Sci.* **18**, 125-127.

Michaud, E. J., Bultman, S. J., Stubbs, L. J., and Woychik, R. P.(1993). The embryonic lethality of homozygous lethal yellow mice ($A^y/A^y$) is associated with the disruption of a novel RNA-binding protein. *Genes & Dev.* **7**, 1203-1213.

Michaud, J. L., Lapointe, F., and Le Douarin, N. M.(1997). The dorsoventral polarity of the presumptive limb is determined by signals produced by the somites and by the lateral somatopleure. *Development* **124**, 1453-1463.

Michel, U., Farnworth, P., and Findlay, J. K.(1993). Follistatins: more than follicle-stimulating hormone suppressing proteins. *Mol. Cell. Endocrinol.* **91**, 1-11.

Miller, J., McLachlan, A. D., and Klug, A.(1985). Repetitive zinc-binding domains in the protein transcription factor IIIA from Xenopus oocytes. *EMBO J.* **4**, 1609-1614.

Miller, M. W., Duhl, D. M. J., Vrieling, H., Cordes, S. P., Ollmann, M. M., Winkes, B. M., and Barsh, B. S.(1993). Cloning of the mouse *agouti* gene predicts a secreted protein ubiquitously expressed in mice carrying the *lethal yellow* mutation. *Genes & Dev.* **7**, 454-467.

Mo, R., Freer, A. M., Zinyk, D. L., Crackower, M. A., Michaud, J., Heng, H. H. -Q., Chik, K. W., Shi, X. -M., Tsui, L. -C., Cheng, S. H., Joyner, A. L., and Hui, C. -c.(1997). Specific and redundant functions of *Gli2* and *Gli3* zinc finger genes in skeletal patterning and development. *Development* **124**, 113-123.

Monaco, A. P. and Larin, Z.(1994). YACs, BACs, PACs and MACs: artificial chromosomes as research tools. *Trends Biotech.* **12**, 280-286.

Montgomery, J. C., Guarnieri, M. H., Tartaglia, K. E., and Flaherty, L. A.(1994). High-resolution genetic map and YAC contig around the mouse neurological locus reeler. *Mamm. Genome* **5**, 756-761.

Motzny, C. K. and Holmgren, R.(1995). The *Drosophila* Cubitus interruptus protein and its role in the *wingless* and *hedgehog* signal transduction pathways. *Mech. Dev.* **52**, 137-150.

Muenke, M. and Schell, U.(1995). Fibroblast-growth-factor receptor mutations in human skeletal disorders. *Trends Genet.* **11**, 308-313.

Muragaki, Y., Mundlos, S., Upton, J., and Olsen, B. R.(1996). Altered growth and branching patterns in synpolydactyly caused by mutations in HOXD13. *Science* **272**, 548-551.

Nadeau, J. H.(1991). *Advanced techniques in chromosome research* (Adolph, K., Ed.) Marcel Dekker, New York, N.Y.. 269-296.

Nadeau, J. H. and Kosowsky, M.(1991). Mouse map of paralogous genes. *Mamm. Genome* **1**, s433-s460.

Nakamura, Y., Wada, K. -N., Wada, Y., Doi, H., Kanaya, S., Gojobori, T., and Ikemura, T.(1996). Codon usage tabulated from the international DNA sequence databases. *Nucleic Acids Res.* **24**, 214-215.

Nakano, Y., Guerrero, I., Hidalgo, A., Taylor, A. M., and Whittle, J. R. S.(1989). The *Drosophila* segment polarity gene *patched* encodes a protein with multiple potential membrane spanning domains. *Nature* **341**, 508-513.

Niswander, L., Jeffrey, S., Martin, G. R., and Tickle, C.(1994). A positive feedback loop coordinates growth and patterning in the vertebrate limb. *Nature* **371**, 609-612.

Niswander, L. and Martin, G. R.(1992). F*gf4* expression during gastrulation, myogenesis, limb and tooth development. *Development* **114**, 755-768.

Niswander, L., Tickle, C., Vogel, A., Booth, I., and Martin, G. R.(1993). FGF4 replaces the apical ectodermal ridge and directs outgrowth and patterning of the limb. *Cell* **75**, 579-587.

No, D., Yao, T. -P., and Evans, R. M.(1996). Ecdysone-inducible gene expression in mammalian cells and transgenic mice. *Proc. Nat. Acad. Sci.* **93**, 3346-3351.

Ochman, H. and Gerber, A. S.(1988). Genetic applications of an inverse polymerase chain reaction. *Genetics* **120**, 621-623.

Oda, S., Nishimatsu, S. -L., Murakami, K., and Ueno, N.(1995). Molecular cloning and functional analysis of a new activin b subunit - A dorsal mesoderm inducing activity in *Xenopus*. *Biochem. Biophys. Res. Comm.* **210**, 581-588.

Ohno, S.(1970). *Evolution by gene duplication*, Springer Verlag, Berlin/Heidelberg/New York.

Orenic, T. V., Slusarski, D. C., Kroll, K. L., and Holmgren, R. A.(1990). Cloning and characterisation of the segment polarity gene *cubitus interuptus Dominant* of *Drosophila*. *Genes & Dev.* **4**, 1053-1067.

Parr, B. A. and McMahon, A. P.(1995). Dorsalising signal *Wnt7a* required for normal polarity of D-V and A-P axes in the mouse limb. *Nature* **374**, 350-353.

Parr, B. A., Shea, M. J., Vassileva, G., and McMahon, A. P.(1993). Mouse *Wnt* genes exhibit discrete domains of expression in the early embryonic CNS and limb buds. *Development* **119**, 247-261.

Parrish, J. E. and Nelson, D. L.(1993). Methods for finding genes: a major rate limiting step in positional cloning. *Gen. Anal. Tech. Appl.* **10**, 29-41.

Patel, K., Sheer, D., and Hampton, D. M.(1993). Junction trapping - a simple method for the isolation of YAC-insert termini. *Genetic analysis-Techniques and applications* **10**, 42-48.

Pavletich, N. P. and Pabo, C. O.(1993). Crystal structure of a five-finger GLI-DNA complex: New perspectives on zinc fingers. *Science* **261**, 1701-1707.

Peters, J. M.(1994). Proteasomes - Protein degradation machines of the cell. *Trends Biochem. Sci.* **19**, 377-382.

Peters, JM., Cejka, Z., Harris, J. R., Kleinschmidt, J. A., and Baumeister, W.(1993). Structural features of the 26S proteasome complex. *J. Mol. Biol.* **234**, 932-937.

Peterson, K. R., Clegg, C. H., Li, Q., and Stamatoyannopoulos, G.(1997). Production of transgenic mice with yeast artificial chromosomes. *Trends Genet.* **13**, 61-66.

Pohl, T. M., Mattei, M. -G., and Ruther, U.(1990). Evidence for allelism of the recessive insertional mutation *add* and the dominant mouse mutation *extra-toes* (*Xt*). *Development* **110**, 1153-1157.

Ragoussis, J.(1995). *Methods in Molecular Biology, Vol 54: YAC Protocols* (Markie, D., Ed.) Humana Press Inc., Totowa N.J.. 69-74.

Rancourt, D. E., Tzuzuki, T., and Capecchi, M. R.(1995). Genetic interaction between *Hoxb5* and *Hoxb6* is revealed by non-allelic noncomplementation. *Genes & Dev.* **9**, 108-122.

Reeves, R. H., Crowley, M. R., O'Hara, B. F., and Gearhart, J. D.(1990). Sex, strain and species differences affect recombination across an evolutionarily conserved fragment of mouse chromosome16. *Genomics* **8**, 141-148.

Riddle, R. D., Ensini, M., Nelson, C., Tsuchida, T., Jessel, T. M, and Tabin, C.(1995). Induction of the LIM homeobox gene *Lmx1* by WNT7a establishes dorsoventral pattern in the vertebrate limb. *Cell* **83**, 631-640.

Riddle, R. D., Johnson, R. L., Laufer, E., and Tabin, C.(1993). *Sonic hedgehog* mediates the polarizing activity of the ZPA. *Cell* **75**, 1401-1416.

Riley, J., Butler, R., Ogilvie, R., Finniear, R., Jenner, D., Powell, S., Anand, R., Smith, J. C., and Markham, A. F.(1990). A novel, rapid method for the isolation of terminal sequences from yeast artificial chromosome (YAC) clones. *Nucleic Acids Res.* **18**, 2887-2890.

Rinchik, E. M.(1991). Chemical mutagenesis and fine structure functional analysis of the mouse genome. *Trends Genet.* **7**, 15-21.

Rinchik, E. M., Carpenter, D. A., and Selby, P. B.(1990). A strategy for fine structure functional analysis of a 6- to 11- centimorgan region of mouse chromosome 7 by high-efficiency mutagenesis. *Proc. Nat. Acad. Sci.* **87**, 896-900.

Ritvos, O., Tuuri, T., Eramaa, M., Sainio, K., Hilden, K., Saxen, L., and Gilbert, S. F.(1995). Activin disrupts epithelial branching morphogenesis in developing glandular organs of the mouse. *Mech. Dev.* **50**, 229-245.

Robbins, D. J., Nybakken, K. E., Kobayashi, R., Sisson, J. C., Bishop, J. M., and Therond, P. P.(1997). Hedgehog elicits signal transduction by means of a large complex containing the kinesin-related protein Costal2. *Cell* **90**, 225-234.

Robert, B., Lyons, G., Simandl, B. K., Kuroiwa, A., and Buckingham, M.(1991). The apical ectodermal ridge regulates *Hox-7* and *Hox-8* gene expression in developing chick limb buds. *Genes & Dev.* **5**, 2363-2374.

Robert, B., Montagutelli, X., Houzelstein, D., Ferland, L., Cohen, A., Buckingham, M., and Guenet, J-L.(1994). *Msx1* is close but not allelic to either *Hm* or *Hx* on mouse chromosome 5. *Mamm. Genome* **5**, 446-449.

Robert, B., Sassoon, D., Jacq, B., Gehring, W., and Buckingham, M.(1989). *Hox-7*, a mouse homeobox gene with a novel pattern of expression during mouse embryogenesis. *EMBO J.* **8**, 91-100.

Roberts, C. W. M., Shutter, J. R., and Korsmeyer, S. J.(1994). *Hox11* controls the genesis of the spleen. *Nature* **368**, 747-749.

Roberts, C. W. M., Sonder, A. M., Lumsden, A., and Korsmeyer, S. J.(1995). Developmental expression of *Hox11* and specification of splenic cell fate. *Am. J. Pathol.* **146**, 1089-1101.

Roberts, W. M., Douglass, E. C., Peiper, S. C., Houghton, P. J., and Look, A. T.(1989). Amplification of the gli gene in childhood sarcomas. *Cancer Research* **49**, 5407-5413.

Rock, K. L., Gramm, C., Rothstein, L., Clark, K., Stein, R., Dick, L., Hwang, D., and Goldberg, A. L.(1994). Inhibitors of the proteasome block the degradation of most cell proteins and the generation of peptides presented on MHC class I molecules. *Cell* **78**, 761-771.

Rodriguez-Esteban, C., Schwabe, J. W. R., De La Pena, J., Foys, B., Eshelman, B., and Belmonte, J. C. I.(1997). *Radical fringe* positions the apical ectodermal ridge at the dorsoventral boundary of the vertebrate limb. *Nature* **386**, 360-366.

Roessler, E., Belloni, E., Gaudenz, K., Jay, P., Berta, P., Scherer, S. W., Tsui, L. C., and Muenke, M.(1996). Mutations in the human *Sonic hedgehog* gene cause holoprosencephaly. *Nature Genet.* **14**, 357-360.

Rooze, M. A.(1977). *Morphogenesis and malformation of the limb* (Bergsman, D. and Lenz, W., Eds.) Liss, New York. 69-95.

Rosenberg, U. B., Schroder, C., Preiss, A., Kienlin, A., Cote, S., Riede, E., and Jackle, H.(1986). Structural homology of the product of the *Drosophila Kruppel* gene with *Xenopus* transcription factor IIIA. *Nature* **319**, 336-339.

Rowe, L. B., Nadeau, J. H., Turner, R., Frankel, W. N., Letts, V. A., Eppig, J. T., Ko, M. S. H., Thurston, S. J., and Birkenmeier, E. H.(1994). Maps from two interspecific backcross DNA panels available as a community genetic mapping resource. *Mamm. Genome* **5**, 253-274.

Southern, E. M.(1975). Detection of specific sequences amoung DNA fragments separated by gel electrophoresis. *J. Mol. Biol.* **98**, 503-517.

Stallings, R. L., Ford, A. F., Nelson, D., Torney, D. C., Hilderbrand, C. E., and Moyzis, R. K.(1991). Evolution and distribution of (GT)n repetitive sequences in mammalian genomes. *Genomics* **10**, 807-815.

Stone, D. M., Hynes, M., Armanini, M., Swanson, T. A., Gu, Q., Johnson, R. L., Scott, M. P., Pennica, D., goddard, A., Phillips, H., Noll, M., Hooper, J. E., Desauvage, F., and Rosenthal, A.(1996). The tumour suppressor gene *patched* encodes a candidate receptor for *Sonic hedgehog*. *Nature* **384**, 129-134.

Storm, E. E., Huynh, T. V., Copeland, N. G., Jenkins, N. A., Kingsley, D. M., and Lee, S. J.(1994). Limb alterations in *brachypodism* mice due to mutations in a new member of the TGF-b-superfamily. *Nature* **368**, 639-643.

Summerbell, D.(1974). A quantitative analysis of the effect of excision of the AER from the chick limb bud. *J. Embryol. Exp. Morph.* **32**, 651-660.

Summerbell, D.(1983). The effect of local application of retinoic acid to the anterior margin of the developing chick limb bud. *J. Embryol. Exp. Morph.* **78**, 269-289.

Summerbell, D., Lewis, J., and Wolpert, L.(1973). Positional information in chick limb morphogenesis. *Nature* **224**, 492-496.

Swaffield, J. C., Melcher, K., and Johnston, S. A.(1995). A highly conserved ATPase protein as a mediator between acidic activation domains and the TATA-binding protein. *Nature* **374**, 88-91.

Sweet, H. O.(1982). *Hm* and *Hx* are not alleles. *Mouse News Lett.* **66**, 66.

Tamura, T., Nagy, I., Lupas, A., Lottspeich, F., Cekja, Z., Schoofs, G., Tanaka, K., Demot, R., and Baumeister, W.(1995). The first characterisation of a eubacterial proteasome-the 20S complex of *Rhodococcus*. *Curr. Biol.* **5**, 766-774.

Taylor, B. A.(1989). *Genetic variants and strains of the laboratory mouse* (Lyon, M. F. and Searle, A. G., Eds.) 2nd edition. Oxford University Press, Oxford. 773-796.

Tickle, C.(1981). The number of polarizing region cells required to specify additional digits in the developing chick wing. *Nature* **289**, 295-298.

Tickle, C., Alberts, B., Wolpert, L., and Lee, J.(1982). Local application of retinoic acid to the limb bud mimics the action of the polarizing region. *Nature* **296**, 564-566.

Tickle, C. and Eichele, G.(1994). Vertebrate limb development. *Ann. Rev. Cell Biol.* **10**, 121-152.

Tickle, C., Lee, J., and Eichele, G.(1985). A quantitative analysis of the effect of all-*trans*-retinoic acid on the pattern of chick wing development. *Dev. Biol.* **109**, 82-95.

Tommerup, N. and Nielsen, F.(1983). A familial reciprocal translocation t(3;7)(p21.1;p13) associated with the Greig polysyndactyly-craniofacial anomalies syndrome. *Am. J. Med. Gen.* **16**, 313-321.

Trask, B. J., Allen, S., Massa, H., Fertitta, A., Sachs, R., van den Engh, G., and Wu, M.(1993). Studies of metaphase and interphase chromosomes using fluorescence *in situ* hybridisation. *Cold Spring Harbour Symp. Quant. Biol.* **58**, 767-775.

Trask, B. J., Massa, H., Kenwrick, S., and Gitschier, J.(1991). Mapping of human chromosome Xq28 by two-colour fluorescence in situ hybridisation of DNA sequences to interphase cell nuclei. *Am. J. Hum. Genet.* **48**, 1-15.

Trask, B. J., Pinkel, D., and van den Engh, G.(1989). The proximity of DNA sequences in interphase cell nuclei is correlated to genomic distance and permits ordering of cosmids spanning 250 kilobase pairs. *Genomics* **5**, 710-717.

Trent, C., Wood, W. B., and Hovitz, H. R.(1988). A novel dominant transformer allele of the sex-determining gene *her-1* of *Caenorhabditis elegans*. *Genetics* **120**, 145-157.

Tsukurov, O., Boehmer, A., Flynn, J., Nicolai, J. -P., Hamel, B. C. J., Traill, S., Zaleske, D., Mankin, H. J., Yeon, H., Ho, C., Tabin, C., Seidman, J. G., and Seidman, C.(1994). A complex bilateral polysyndactyly disease locus maps to chromosome 7q36. *Nature Genet.* **6**, 282-286.

Tykocinski, M. and Max, E. E.(1984). CG nucleotide clusters in MHC genes and in 5' demethylated genes. *Nucleic Acids Res.* **12**, 4385-4396.

Valdes, J. M., Tagle, D. A., and Collins, F. S.(1994). Island rescue PCR: A rapid and efficient method for isolating transcribed sequences from yeast artificial chromosomes and cosmids. *Proc. Nat. Acad. Sci.* **91**, 5377-5381.

Vale, W., Hsueh, A., Rivier, C., and Yu, J.(1990). *Peptide growth factors and their receptors* (Sporn, M. and Roberts, A., Eds.) Springer-Verlag, Berlin. 211-248.

van den Eijnden-van Raaij, A. J. M., Feijen, A., Lawson, K. A., and Mummery, C. L.(1992). Differential expression of inhibin subunits and follistatin, but not of activin receptor type II, during early murine embryonic development. *Dev. Biol.* **154**, 356-365.

van den Heuvel, M. and Inghan, P. W.(1996). *smoothened* encodes a receptor-like serpentine protein required for hedgehog signalling. *Nature* **382**, 547-551.

van der Hoeven, F., Schimmang, T., Vortkamp, A., and Ruther, U.(1993). Molecular linkage of the morphogenetic mutation *add* and the zinc finger gene *Gli3*. *Mamm. Genome* **4**, 276-277.

Vassalli, A., Matzuk, M. M., Gardner, H. A. R., Lee, K. -F., and Jaenisch, R.(1994). Activin/inhibin bB subunit gene disruption leads to defects in eyelid development and female reproduction. *Genes & Dev.* **8**, 414-427.

Vitaterna, M. H., King, D. P., Chang, A. M., Kornhauser, J. M., Lowery, P. L., McDonald, J. P., Dove, W. F., Pinto, L. H., Turek, F. W., and Takahashi, J. S.(1994). Mutagenesis and mapping of a mouse gene Clock, essential for circadian behaviour. *Science* **264**, 719-725.

Vogel, A., Rodruiguez, C., Warnken, W., and Izpisua-Belmonte, J. C.(1995). Dorsal cell fate specified by chick *Lmx1* during vertebrate limb development. *Nature* **378**, 716-720.

Vogel, A. and Tickle, C.(1993). FGF4 maintains polarizing activity of posterior limb bud cells in vivo and in vitro. *Development* **119**, 199-206.

Vortkamp, A., Franz, T., Gessler, M., and Grzeschik, K.(1992). Deletion of GLI3 supports the homology of the human Greig cephalopolysyndactyly syndrome (GCPS) and the mouse mutant extra toes (Xt). *Mamm. Genome* **3**, 461-463.

Vortkamp, A., Gessler, M., and Grzeschik, K. -H.(1991). GLI3 zinc-finger gene interrupted by translocations in Greig syndrome families. *Nature* **352**, 539-540.

Vortkamp, A., Gessler, M., and Grzeschik, K. -H.(1995). Identification of optimized target sequences for the GLI3 zinc finger protein. *DNA and Cell Biology* **14**, 629-634.

Vortkamp, A., Gessler, M., Le Paslier, D., Elaswarapu, R., Smith, S., and Grzeschik, K. -H.(1994). Isolation of a yeast artificial chromosome contig spanning the Greig cephalopolysyndactyly syndrome (GCPS) gene region. *Genomics* **22**, 563-568.

Vortkamp, A., Heid, C., Gessler, M., and Grzeschik, K. -H.(1995). Isolation and characterisation of a cosmid contig for the GCPS gene region. *Hum. Genet.* **95**, 82-88.

Wagner, K., Kroisel, P. M., and Rosenkranz, W.(1990). Molecular and cytogenetic analysis in two patients with microdeletions of 7p and Greig syndrome: Hemizygosity for PGAM2 and TCRG genes. *Genomics* **8**, 487-491.

Walterhouse, D., Ahmed, M., Slusarski, D., Kalamaras, J., Boucher, D., Holmgren, R., and Innaccone, P.(1993). *gli*, a zinc finger transcription factor and oncogene, is expressed during normal mouse development. *Dev. Dynamics* **196**, 91-102.

Waxman, L., Fagan, J. M., and Goldberg, A. L.(1987). Demonstration of two distinct high molecular weight proteases in rabbit reticulocytes, one of which degrades ubiquitin conjugates. . *J. Biol. Chem.* **262**, 2451-2457.

Weber, J. L. and May, P. E.(1989). Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. *Am. J. Hum. Genet.* **44**, 388-396.

Weil, D., Blanchard, S., Kaplan, J., Guilford, P., Walsh, J., Mburu, P., Varela, A., Levillers, J., Weston, M. D., Kelly, P. M., Kimberling, W. J., Wagenaar, M., Leviacobas, F., Largetpiet, D., Munnich, A., Steel, K. P., Brown, S. D. M., and Petit, C.(1995). Defective myosin VIIA gene responsible for Usher syndrome type-1B. *Nature* **374**, 60-61.

Wenzel, T. and Baumeister, W.(1995). Conformational constraints in protein degradation by the 20S proteasome. *Nature Struct. Biol.* **2**, 199-203.

Wild, A., Kalff-Suske, M., Vortkamp, A., Bornholdt, D., Konig, R., and Grzeschik, K.-H. (1997). Point mutations in human *GLI3* cause Greig syndrome. *Hum. Mol. Genet.* **6**, 1979-1984.

Wilkie, A. O. M.(1994). The molecular basis of genetic dominance. *J. Med. Genet.* **31**, 89-98.

Williamson, P., Holt, S., Townsend, S., and Boyd, Y.(1995). A somatic cell hybrid panel for mouse gene mapping characterised by PCR and FISH. *Mamm. Genome* **6**, 429-432.

Winship, P. R.(1989). An improved method for directly sequencing PCR amplified material using dimethyl sulphoxide. *Nucleic Acids Res.* **17**, 1266.

Winter, R. M. and Huson, S. M.(1988). Greig cephalopolysyndactyly syndrome: A possible mouse homologue (Xt-Extra Toes). *Am. J. Med. Gen.* **31**, 793-798.

Winter, R. M. and Tickle, C.(1993). Syndactylies and polydactylies: Embryological overview and suggested classification. *Eur. J. Hum. Genet.* **1**, 96-104.

Wolpert, L.(1971). Positional information and pattern formation. *Curr. Top. Dev. Biol.* **6**, 183-224.

Wong, A. J., Bigner, S. H., Bigner, D. D., Kinzler, K. W., Hamilton, S. R., and Vogelstein, B.(1987). Increased expression of the epidermal growth factor receptor gene in malignant gliomas is invariably associated with gene amplification. *Proc. Nat. Acad. Sci.* **84**, 6899-6903.

Yang, Y. Z. and Niswander, L.(1995). Interaction between the signalling molecules *Wnt7a* and *Shh* during vertebrate limb development - dorsal signals regulate anteroposterior patterning. *Cell* **80**, 939-947.

Yokota, H., Vandenengh, G., Hearst, J. E., Sachs, R. K., and Trask, B. J.(1995). Evidence for the organisation of chromatin in megabase pair-sized loops arranged along a random walk path in the human $G_0/G_1$ interphase nucleus. *J. Cell Biol.* **130**, 1239-1249.

Zakeri, Z., Quaglino, D., and Singh Ahuja, H.(1994). Apoptotic cell death in the mouse limb and its suppression in the Hammertoe mutant. *Dev. Biol.* **165**, 294-297.

Zarkower, D. and Hodgkin, J.(1992). Molecular analysis of the *C. elegans* sex-determining gene *tra-1*: A gene encoding two zinc finger proteins. *Cell* **70**, 237-249.

Zarkower, D. and Hodgkin, J.(1993). Zinc fingers in sex determination: only one of the two *C. elegans* Tra-1 proteins binds DNA *in vitro*. *Nucleic Acids Res.* **21**, 3691-3698.

Zimmer, E. A., Martin, S. L., Beverly, S. M., Kan, Y. W., and Wilson, A. C.(1980). Rapid duplication and loss of genes coding for the alpha chains of haemoglobin. *Proc. Nat. Acad. Sci.* **7**, 2158-2162.

Zwilling, E.(1956). Interaction between limb bud ectoderm and mesoderm in the chick embryo. *J. Exp. Zool.* **132**, 157-72.

# APPENDIX

## Publications presented during the course of this thesis

Mammalian Genome 8, 352–354 (1997).

Mammalian Genome



# Expression and mapping of the mouse S7/*Pmsc2* gene, homolog of an essential mitotic gene in yeast

**Jeremy Allen,* Laurence Colleaux,** Duncan Davidson, Elisabeth Graham, Muriel Lee, Robert Hill, Cathy Abbott,*** Colin Gordon**

MRC Human Genetics Unit, Western General Hospital, Crewe Road, Edinburgh, EH4 2XU, Scotland, UK

Received: 13 August 1996 / Accepted: 23 December 1996

Polyubiquitinated proteins are degraded by a 26S ATP-dependent protease. This large complex contains a proteolytic core, previously identified as the 20S proteasome, which associates with regulatory components to form the 26S particle (reviewed by Rechsteiner et al. 1993; Hershko and Ciechanover 1992). Recent studies in yeast show that mutations in two of the regulatory components, S4 and S7, interrupt mitosis, indicating a role for the proteasome in cell cycle regulation (Gordon et al. 1993; Ghislain et al. 1993). In the *Drosophila* embryo, a cell-specific accumulation of the proteasome suggests that this complex regulates developmental processes correlated with cell proliferation and morphogenesis (Klein et al. 1990). To investigate a possible role for the proteasome in mammalian development, we used the previously isolated mouse cDNA, originally designated *MSS1* (Gordon et al. 1993), homologous to the yeast and human (Dubiel et al. 1993) S7 regulatory components. The mouse S7 homolog has recently been given the locus symbol *Pmsc2*. We report the expression pattern of the *Pmsc2* gene in mouse embryos and the map location in the mouse genome.

Northern blot analysis shows that all the adult tissues examined express *Pmsc2* but at variable levels, and in embryos there is a slight increase in expression until day E14.5, followed by a decrease by E16.5 (data not shown). There is a single mRNA species of approximately 1.5 kb in all expressing tissues. Expression in a broad range of cell types would be expected for the proteasome complex, which may have a central role in normal cellular functions. The variability in expression, however, could indicate that cells which sustain a high metabolic activity, perhaps those cells which are rapidly proliferating or undergoing differentiation, may have higher levels of the complex.

To investigate the spatial distribution of the 26S subunit in embryos *in situ* hybridizations were performed on whole embryos (whole mount *in situ* hybridization method described by Wilkinson 1992), with a *Pmsc2* antisense probe. Hybridization probes were generated from the 1.4-kb S7 cDNA cloned (pBluescript II KS [Stratagene]) such that antisense is generated from the T7 promoter and sense from the T3 promoter. (The sense probe showed only low, uniform labeling [Fig. 1A]). By E11 (Fig. 1B) of development (E0.5 being the afternoon of the day of vaginal plug detection), *Pmsc2* expression is detectable at specific sites. In both forelimb and hindlimb buds, the *Pmsc2* gene expression is restricted distally (Fig. 1B,C). The highest levels are detected in the mesenchymal cells at the distal rim of the bud under the apical ectodermal ridge (AER) corresponding to the progress zone. No strong hybridization signal is present in the AER itself (Fig. 1D). In E12.5 embryos (Fig. 1E), *Pmsc2* is expressed in the mesenchyme lining the digits but not throughout the whole of the interdigital region. The majority of the interdigital mesenchyme undergoes programmed cell death; thus, *Pmsc2* expression is not strictly correlated with this morphogenetic process.

An appreciable number of classical mouse limb mutations are known (Hill and Townley 1996) whose phenotypes may relate to abnormal cell proliferation or morphogenesis within the progress zone. To investigate a possible relationship between one of these mutations and the *Pmsc2* gene, we determined the genetic map position of *Pmsc2*. We initially analyzed an interspecific backcross [(C57BL/6J × Spret/Ei)$F_1$ × Spret/Ei] panel of 94 animals (The Jackson Laboratory Interspecific Backcross Panel 2; Rowe et al. 1994). An RFLV that distinguished the *Mus spretus* and the C57BL/6 *Pmsc2* alleles was found. An intronic fragment generated by PCR, with oligonucleotide primers (primer 1 5′-GGAGCTCGAATGGTTCGTGAG-3′ and primer 2 5′-TACTGTTTGGACACAGGCGGGCC-3′) predicted from the *Pmsc2* coding region, was variant for *Sac*I digestion (using Boehringer *Sac*I according to manufacturer's instructions). The 1-kb PCR product generated from the C57BL/*6J* allele digested with *Sac*I yielded a 350-bp and a 650-bp fragment; the *Mus spretus* allele does not contain a *Sac*I site. All PCR assays were performed with AmpliTaq in the standard buffers supplied by the manufacturer (Perkin-Elmer). The conditions for these polymerase chain reactions consisted of an initial step to melt the mouse genomic DNA at 94°C for 4 min. The number of cycles and the annealing temperature varied according to the primers used. The assay products were fractionated on 1–4% (NuSieve GTG) agarose gel. For *Pmsc2* primers, the reaction was amplified for 40 cycles, each cycle consisting of the following incubation conditions: 94°C for 45 s, 61°C annealing temperature for 45 s, and 72°C extension temperature for 1 min.

The strain distribution pattern (SDP) was determined and analyzed by Map Manager (Manly 1993; data on the World Wide Web at http://www.jax.org/resources/documents/cmdata), which indicated that *Pmsc2* was concordant with the *Xmv-45, Htr5a,* and *Pep1b* loci located on proximal Chr 5. This placed *Pmsc2* in a region containing two limb mutations, hemimelic extra toes (*Hx*) and hammer toe (*Hm*).

*Hm* and *Hx* are tightly linked limb mutations that may be allelic (Green 1989). In the heterozygote the *Hx* phenotype is characterized by preaxial polydactyly on all four feet and tibial and radial hemimelia. The basis for this phenotype has been related to defects within the progress zone of the developing limb (Knudsen and Kochhar 1981). In contrast, the *Hm* mutation affects the limb at a later stage and interferes with the process of interdigital mes-

*Correspondence to:* R. Hill

**Present address:* Dept. of Neurobiology, The Babraham Institute, Babraham, Cambridge CB2 4AT, UK.

***Present address:* Faculte de Medecine, 27 Boulevard Jean Moulin, 13385 Marseille CEDEX 5, France.

****Present address:* Human Genetics Unit, Dept of Medicine, University of Edinburgh, Molecular Medicine Centre, Western General Hospital, Crewe Rd, Edinburgh, UK.

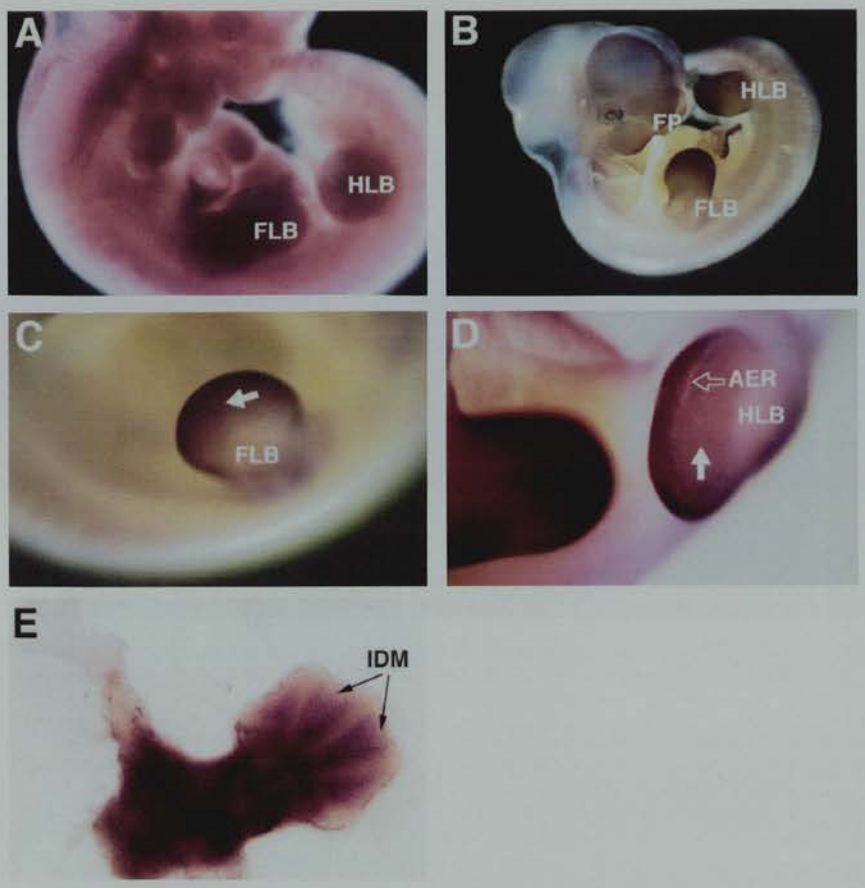



**Fig. 1.** Localization of the *Pmsc2* transcript by whole-mount in situ hybridisation. (A,B,C, and D) E11 and (E) E12.5 mouse embryos were hybridized with either a sense (A) or antisense (B,C,D, and E) *Pmsc2* riboprobe. **A** shows control hybridization with the sense probe. Each control was stained two to three times longer than experimental embryos, and no specific accumulation of hybridization was detected. **B** shows overall hybridization of the *Pmsc2* probe with high accumulation in both the forelimbs (FLB) and the hindlimbs (HLB) and the frontal nasal processes (FP). **C** is higher magnification of the forelimb showing predominant hybridization at the distal tip of the limb (indicated by the closed arrow), and **D** shows a stripe of nonhybridizing distal limb cells which relates to the apical ectodermal ridge (AER) (open arrow). Closed arrow shows hybridization to distal limb mesenchyme as in C. In older embryos, E12.5, (**E**) expression of *Pmsc2* accumulates in the interdigital mesenchyme (IDM), particularly in mesenchyme lining the digits (digits are areas of lightly labeled hybridization).

enchymal cell death (Zakeri et al. 1994), resulting in webbed and flexed digits. Since the expression pattern of the *Pmsc2* gene is found in both the progress zone and later, at the margins of the interdigital mesenchyme, thus relating to both these phenotypes, we extended the mapping analysis to further investigate the relationship between *Pmsc2* and these two mouse mutations.

*Emv1* maps approximately 1 cM distal of *Hx*, which is flanked on the proximal side by *En2* (Martin et al. 1990). Thus, to investigate possible linkage of *Pmsc2* with *Hx*, we mapped *Pmsc2* relative to the closely linked *En2* gene on the Interspecific Backcross Panel. A *Mae*II restriction site variation in the 3′ untranslated region of *En2* was used to determine the SDP. The *En2* 3′ untranslated region was amplified by PCR (primer 1 5′-CTCCCGAATCTAGAGGACATGTG-3′ and primer 2 5′-CACTAAACAGAAGACATTGTGGAG-3′) to yield a 232-bp fragment. The *Mus spretus* allele contained the *Mae*II restriction site, which upon digestion gives a 189-bp and a 43-bp fragment. The PCR assay was done as above at an annealing temperature of 65°C. In this backcross *En2* and *Pmsc2* cosegregate, suggesting close linkage with the chromosomal region containing the two limb mutations.

To refine the map position, we used the European Collaborative Interspecific Backcross, EUCIB, which provides an average genetic resolution of 0.1 cM. From the total of 984 backcross progeny, a panel of 384 mice were selected. In this backcross *D5Nds8* was the most proximal anchor marker on Chr 5; however, it is distal to *En-2* and thus *Hm* and *Hx*. An additional proximal marker, *D5Mit226* microsatellite, was mapped on this panel (PCR conditions as above, except that an annealing temperature of 55°C was used). The 55 mice that were recombinant between this marker and *D5Nds8* were selected for further analysis. In this screen, double recombinations will have been missed; however, these would not add significantly to this analysis. A distribution pattern for the *Pmsc-2, En2,* and *Emv1* alleles for these 55 mice was determined (Fig. 2). The analyses of the RFLVs for *Pmsc2* and *En2* were as above. The *Emv1* locus was mapped with a closely associated microsatellite repeat isolated from a genomic clone containing *Emv1*. The PCR assay (primers 5′-CTCAGTGTTGTCCTGATACGTGAC-3′ and 5′CCATTCCT-CACGGAGCCAAGTG-3′) generated a 100-bp fragment for the C57BL/6 allele and a 120-bp fragment for the *Mus spretus* allele. (PCR conditions, with an annealing temperature of 65°C, were as above). Of the 55 recombinant mice, six did not type for all of the markers (Fig. 2). The predicted order and relative distance of these markers (Fig. 2) are as follows-

*D5Mit266*–4.2 ± 1.0 cM–*Pmsc2*–4.0 ± 1.0 cM–*En2*–1.5 ± 0.6 cM–*Emv1*–3.2 ± 0.9 cM–*D5Nds8*

This places *Pmsc2* on the proximal side of *En-2* outside the region mapped for *Hm* and *Hx* (Martin et al. 1990).

FISH with mouse genomic clones for *En2, S7/Pmsc2,* and *Emv1* (λ clones from a λFIXII library [Stratagene]) confirmed the order of the genes (Fig. 3). To establish the relative position of the

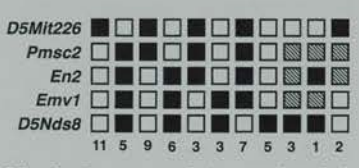



**Fig. 2.** Summary of the results of the EUCIB interspecific backcross analysis. Genes mapped in the analysis are shown to the left. Each column represents the chromosome identified in the $F_2$ progeny that was inherited from the (C57BL/6 × *M. spretus*) $F_1$ parent. The black boxes represent the C57BL/6 allele, and the white boxes represent the presence of a *M. spretus* allele. Alleles that were not defined are shown as crosshatched boxes. The number of each type of chromosome identified is shown at the bottom.

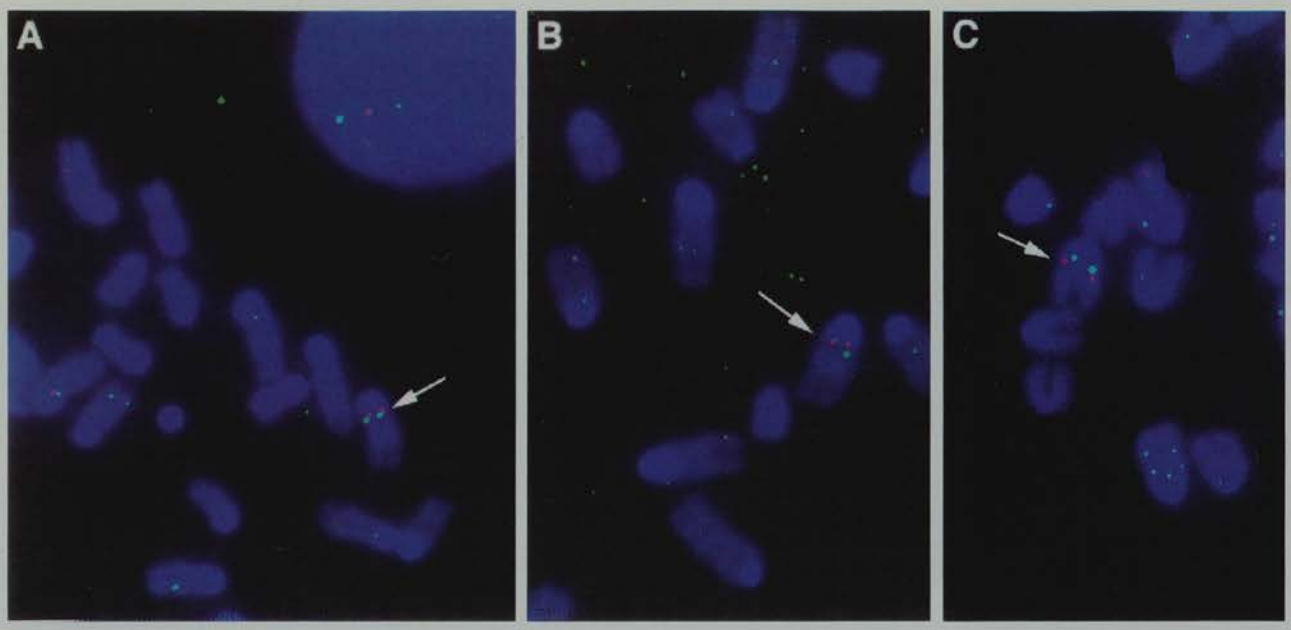



**Fig. 3.** Relative order of *Pmsc2*, *En1*, and *Emv1* by FISH analysis on Chr 5. Panel A shows a typical hybridization with probes for *Emv1* (blue) and *En2* (red); Panel B for *Emv1* (blue) and *Pmsc2* (red); and Panel C for *En2* (red) and *Pmsc2* (blue). For each pair of probes at least 20 Chr 5's were scored.

markers, we performed hybridizations (Fantes et al. 1995) with pairs of probes labeled with either biotin or digoxygenin; biotin was detected with FITC (green signal) and digoxygenin with TR (red signal). The predicted order was *Pmsc2* proximal, *Emv1* distal, with *En2* located between the two genes.

*Pmsc2* is an integral component of the 26S proteasome. Embryonic expression of *Pmsc2* is increased in specific regions of the embryo, suggesting that in these regions there is a requirement for higher levels of the 26S proteasome complex. The early limb shows expression in the distal mesenchyme relating to the progress zone (a region of rapid cell division and morphogenesis) and later in mesenchyme residing within areas undergoing programmed cell death. Increased proteolytic activity may thus be important in these early processes of limb development. Although *Pmsc2* is not responsible for the *Hm/Hx* phenotypes, the chromosomal location corresponds to a conserved syntenic region in human (Chr 7), which relates to the polysyndactyly syndrome (Tsukurov et al. 1994). In addition, at least 15 different subunits are thought to constitute the regulatory complex of the 26S proteasome, and mutations in any one of these may result in limb or other embryonic abnormalities.

*Acknowledgments.* We thank Norman Davidson, Sandy Bruce, and Douglas Stuart for excellent photography. L. Colleaux is supported by the CNRS.

## References

Dubiel W, Ferrell K, Rechsteiner M (1993) Peptide sequencing identifies a modulator of HIV Tat-mediated transactivation, as subunit 7 of the 26S protease. FEBS Lett 323, 276–278

Fantes J, Oghene K, Boyle S, Danes S, Fletcher J, Bruford E, Williamson K, Seawright A, Schedl A, Hanson I, Zehtner G, Bhogal R, Lehrach H, Gregory S, Williams J, Little P, Sellar G, Hoovers J, Mannens M, Weissbach J, Jumien C, van Heyningen V, Bickmore W (1995) A high resolution integrated physical, cytogenetic and genetic map of human chromosome 11: distal p13 to proximal p15.1. Genomics 25, 447–461

Ghislain M, Udvardy A, Mann C (1993) *S. cerevisiae* 26S protease mutants arrest in G2/metaphase. Nature 366, 358–362

Gordon C, McGurk G, Dillon P, Rosen C, Hastie, ND (1993) Defective mitosis due to a mutation in the 26S protease subunit. Nature 366, 355–357

Green M (1989) Catalog of mutant genes and polymorphic loci. In *Genetic Variants and Strains of the Laboratory Mouse,* M Lyon, A Searle eds. (Oxford, UK: Oxford University Press), pp 12–403

Hershko A, Ciechanover A (1992) The ubiquitin system for protein degradation. Annu Rev Biochem 61, 761–807

Hill R, Townley D (1996) Mammalian limb development: a genetic perspective. In *Mammalian Development,* P Lonai, ed. (Harwood Academic Publishers) pp 151–171

Klein U, Gernold M, Kloetzel P (1990) Cell-specific accumulation of *Drosophila* proteasomes(MCP) during early development. J Cell Biol 111, 2275–2282

Knudsen TB, Kochhar DM (1981) The role of morphogenetic cell death during abnormal limb-bud outgrowth in mice heterozygous for the dominant mutation Hemimelia-extra toe. J Embryol Exp Morphol 65 (Suppl.), 289–307

Manly K (1993) A Macintosh program for the storage and analysis of experimental genetic mapping data. Mamm Genome 4, 303–313

Martin G, Richman M, Reinisch S, Nadeau J, Joyner J (1990) Mapping of the two mouse engrailed-like genes—close linkage of *En-1* to Dominant hemimelia (*Dh*) on chromosome 1 and of *En-2* to hemimelic extra-toes (*Hx*) on chromosome 5. Genomics 6, 302–308

Rechsteiner M, Hoffman L, Dubiel W (1993) The multicatalytic and 26S proteases. J Biol Chem 268, 6065–6068

Rowe LB, Nadeau JH, Turner R, Frankel WN, Letts VA, Eppig JT, Ko MSH, Thurston SJ, Birkenmeier EH (1994) Maps from two interspecific backcross DNA panels available as a community genetic mapping resource. Mamm Genome 5, 253–274

Tanaka K, Tamura T, Yoshimura T, Ichihara A (1992) Proteasomes: protein and gene structures. New Biol 4, 173–187

Tsukurov O, Boehmer A, Flynn J, Nicolai J, Hamel B, Traill S, Zaleske D, Mankin H, Yeon H, Ho C, Tabin C, Seidman J, Seidman C (1994) A complex bilateral polysyndactyly disease locus maps to chromosome 7q36. Nature Genet 6, 282–286

Wilkinson DG (1992) Whole mount *in situ* hybridization of vertebrate embryos. In *In Situ Hybridization-A Practical Approach,* DG Wilkinson, ed. (Oxford: Oxford University Press), pp 75–83

Zakeri Z, Quaglino D, Ahuja H (1994) Apoptotic cell-death in the mouse limb and its suppression in the Hammertoe mutant. Dev Biol 165, 294–297

GENOMICS 39, 205–215 (1997)
ARTICLE NO. GE964468

# Cloning and Sequencing of the Mouse *Gli2* Gene: Localization to the *Dominant hemimelia* Critical Region

DAVID C. HUGHES,*,†,[1] JEREMY ALLEN,* GARRY MORLEY,‡ KATE SUTHERLAND,‡ WASIM AHMED,‡ JANE PROSSER,* LAURA LETTICE,* GORDON ALLAN,* MARIE-GENEVIEVE MATTEI,§ MARTIN FARRALL,‡,[2] AND ROBERT E. HILL*

*MRC Human Genetics Unit, Western General Hospital, Crewe Road, Edinburgh EH4 2XU, United Kingdom; †MRC Institute of Hearing Research, University Park, University of Nottingham, Nottingham, NG7 2RD, United Kingdom; ‡MRC Molecular Medicine Group, MRC Clinical Science Centre, Hammersmith Hospital, London RPMS, London, United Kingdom; and §Centre de Genetique Medicale, Hopital d'Enfants de la Timone, Marseille, France

Received July 9, 1996; accepted October 22, 1996

**The *GLI* family of zinc finger genes has been implicated in both neoplastic and developmental disorders. We have cloned and sequenced the mouse homolog of the zinc finger gene *Gli2* and demonstrated significant similarity to the human *GLI3* gene. We have also localized *Gli2* to mouse chromosome 1, in the vicinity of the morphogenetic mutation *Dominant hemimelia* (*Dh*), which is characterized by tibial hemimelia, poly/oligodactyly, and a number of visceral abnormalities, most strikingly absence of the spleen. Using a *Gli2*-associated microsatellite, we demonstrated no recombination between *Dh* and *Gli2* in a *Dh* intraspecific backcross. *Gli2* is expressed in *Dh* heterozygotes and homozygotes. However, using a combination of mismatch analysis and direct sequencing, we have failed to identify any mutations in the coding sequence of *Gli2* from *Dh*. We have also demonstrated that it is unlikely that there are any *Gli* genes in the mouse genome in addition to the previously described *Gli, Gli2,* and *Gli3*.** © 1997 Academic Press

## INTRODUCTION

The human *GLI* gene was originally isolated by virtue of its amplification in a human glioma cell line (Kinzler *et al.,* 1987); two additional members of the human *GLI* family (*GLI2* and *GLI3*) were isolated on the basis of homology in the zinc finger region (Ruppert *et al.,* 1988) and mapped to chromosomes 2 and 7, respectively. Recently it has been demonstrated that *GLI3* is mutated in both the human disorder Grieg cephalopolysyndactlyly syndrome (GCPS) and the mouse *extra-toes* (*Xt*) mutation (Vortkamp *et al.,* 1991, 1992; Hui and Joyner, 1993; Schimmang *et al.,* 1992). In addition, the *Drosophila* segment polarity gene *cubitis interruptus* (*ci*) (Orenic *et al.,* 1990), and the *Caenorhabditis elegans* sex determining gene *tra-1* (Zarkower and Hodgkin, 1992) contain homologous zinc finger domains. We have now isolated and sequenced full-length cDNA clones for mouse *Gli2* and mapped it to a region of mouse chromosome 1 that includes the semidominant mutation *Dominant hemimelia* (*Dh*). *Dh,* which arose spontaneously, is distinguished by tibial hemimelia, preaxial oligo or polydactyly of the hindlimbs, asplenia, a reduced number of ribs and presacral vertebrae, a shortened coelom, a small and defective digestive tract, and abnormalities of the urogenital system, particularly hydropic kidneys in the heterozygous condition (Searle, 1964). Homozygotes display the same range of defects but are more severely affected, frequently dying within a few days of birth (Searle, 1964). However, homozygotes have been reported to survive to maturity and, in the case of one female, to be fertile, demonstrating that the reproductive system is unaffected (Searle, 1964). *Dh* has been previously mapped to chromosome 1 in the vicinity of the homeobox gene *en1* (Higgins *et al.,* 1992; Martin *et al.,* 1990).

## MATERIALS AND METHODS

*cDNA cloning.* An 825-bp DNA probe that encodes 275 aa from the carboxy-terminal portion of the human *GLI3* open reading frame (ORF) was synthesized by RT-PCR from total RNA prepared from a human lymphoblastoid cell line (forward primer, 5′-tac ctg gct cac cag ctc ct-3′; reverse primer, 5′-act cct att gat ttc cgt tg-3′) and [$^{32}$P]CTP radiolabeled using random hexamer priming. This probe was used to isolate cDNA clone Gli2-1 following screening of a LambdaZAPII female mouse lung library that contains both oligo(dT) and random-primed clones (Stratagene). cDNA clones Gli2-2 and

Sequence data from this article have been deposited with the GenBank/EMBL Data Libraries under Accession Nos. X99104 (cDNA) and Y08012 (microsatellite). Note also the following MGD Accession Nos. for the data in this manuscript: MGD-RIEX-212 for the BXD recombinant inbred data; MGD-RIEX-213 for the BXH recombinant inbred data; MGD-CREX-722 for the Dh cross data; MGD-INEX-34 for the *in situ* mapping of Gli2.

[1] To whom correspondence should be addressed at MRC Institute of Hearing Research, University Park, University of Nottingham, Nottingham, NG7 2RD, United Kingdom. Telephone: (115) 925 3425. Fax: (115) 951 8503. E-mail: davidh@ihr.mrc.ac.uk.

[2] Present address: The Wellcome Trust Centre for Human Genetics, Department of Medicine, University of Oxford, Oxford, United Kingdom.



0888-7543/97 $25.00

**TABLE 1**

**Primers Used in PCR Analysis**

| Fragment/primer | Forward primer (a) | Reverse primer (b) |
|---|---|---|
| CA | TTCAGGCAGACCAAAGATAGAACATT | CACTGACATATGTACCATTTTCAT |
| 511/512 | AAGGCCTACTCCCGCCTGGAGAACC | CCTCGTTGGAGTGGGTGCGATTCTG |
| 1 | TGGAGAGTCACCCTTCAGCGCC | AGTTGGGTAGGCATGGTGCTGA |
| 2 | TGGCCTACCAGCAGATCCTGAG | TGTCTTCAGGTTCTCCAGGCGA |
| 3 | CACGAGAGAGCAGAAGCCCTTC | AGGGCTCGCTGCTGCAGGATGA |
| 4 | AGCCACACTGTGGAGGACTGC | GATGGGGTCATAGGAGTCTTGCT |
| 5 | CCATGAGCTCGGCCTACACTGT | GTCGCTGGCCCGCCGTGTGCTG |
| 6 | GGTGCCTGCCCACATCCACTG | GCACAGGCATGTTGCTCTTGTT |
| 7 | CGAGAACTCTAAGCTGCCCAGTC | GTGAGGTTGGTGACACCGCAG |
| 8 | CCTCTGCAGCAATATGGCAGCC | CCCGGTGAAGGCAAAGGCTCAG |
| 9 | CACATGTATGAACAGAATGGAG | GAATGATAGAGGCAAAGGAGCC |
| 10 | CTCGTTCACATGACGCTTGCCA | AAGGCATGTAGACGCAAGTCTT |
| 12 | GATTCGGACCTCTCCCAACTCGCTGG | |
| F187 | CA(T/C)ATIAA(T/C)A(G/A)IGAICA(T/C)ATICA(T/C)GG | |
| E195 | GCICA(A/G)TA(T/C)ATG(C/T)TIGTIGTICA(T/C)ATG | |
| E197 | TGI(T/C)GI(T/C)GCAT(A/G)TGIACIACIA(G/A)CAT | |
| E196 | GGICC(A/G)TGIACIGT(T/C)TTIAC(A/G)TG(T/C)TT | |

*Note.* The sequences of the PCR primers used in this study are listed below (5′ → 3′) according to either the specific name of the primer or the fragment identified.

Gli2-3 were obtained by screening an 8.5-d.p.c. oligo(dT)-primed total mouse embryonic library with a 350-bp fragment derived from clone Gli2-1.

*DNA sequencing and software.* Sanger dideoxy sequencing was performed manually using a commercial kit (Sequenase, USB) by a custom primer strategy using a combination of M13 and plasmid (Bluescript) subclones; the sequences for both strands were determined. Sequences were assembled using the XDAP program and analyzed with the XNIP program; homologies between proteins were graphically assessed with the XSIP program. The GCG package of programs and searches of the EMBL and GenBank DNA databases and NBRF and SwissProt protein databases with the BLAST software package were performed on the computer facilities provided by the MRC HGMP-RC at Hinxton Hall, Cambridge, UK.

*In situ hybridization.* Metaphase spreads were prepared from a WMP male mouse in which all the autosomes except chromosome 19 were in the form of metacentric translocations. *In situ* hybridization was performed as previously described (Mattei *et al.,* 1985) with a 1.2-kb *Pst*I fragment of clone Gli2-1.

*Microsatellite isolation.* A BALB/c mouse genomic lambda library (Clontech) was screened with a 1.2-kb *Pst*I fragment from the *Gli2* cDNA clone Gli2-1, and one clone was identified in $2 \times 10^5$ clones screened. A *Sau*3A "shotgun" library from this lambda clone in Bluescript was hybridized with a $\gamma$-$^{32}$P-labeled $(CA)_{27}$ oligonucleotide. One positive clone was isolated and sequenced. PCR was carried out with primers based on the flanking sequence (CA; Table 1) using AmpliTaq (Perkin–Elmer) according to the manufacturer's instructions for 30–40 cycles of 92°C, 10 s; 60°C, 5 s; 72°C, 30 s. Reactions were electrophoresed on 3% agarose (BRL) gels and stained with ethidium bromide (Sigma).

*RT-PCR.* RT-PCR was performed using a random hexanucleotide primer according to Kawasaki (1990), except 1 $\mu$l of the reverse transcription reaction was used to seed PCR amplifications of 20–50 $\mu$l. The primers used for amplification of *Gli2* are listed in Table 1. Mismatch analysis was performed as previously described (Condie *et al.,* 1993). PCR fragments were sequenced according to the protocol of Winship (1989) with Sequenase (USB).

*Genetic mapping.* DNA samples from the BXD and BXH recombinant inbred strains were obtained from The Jackson Laboratory. DNA was prepared from the *Dh* backcross panel by standard methods (Higgins *et al.,* 1992). MAPpair primers were obtained from Research Genetics, and PCR was performed according to Dietrich *et al.* (1992).

*Fingerprinting analysis.* This procedure was performed essentially as described (Boehm, 1993). cDNA was generated by reverse transcription of 12.5-day mouse embryo RNA (Kawasaki, 1990). The conditions for amplification with primers E195 and E196 were 94°C, 3 min; 35 cycles of 94°C, 45 s; 45°C, 45 s; 72°C, 1 min; and finally 72°C, 5 min. The same conditions were used for primers E187 and E197, except the template source was C57BL/6 DNA (50 ng) and an annealing temperature of 37°C was used for the lower stringency PCR.

## RESULTS

### *cDNA Cloning and Sequencing of Gli2*

As part of a strategy to analyze members of the *GLI* gene family during mammalian development, we screened a mouse lung cDNA library with a human probe derived from the carboxy-terminal portion of the human *GLI3* gene. Approximately $1.5 \times 10^6$ phage were screened, and a single positive clone (Gli2-1) was plaque purified. This clone was restriction mapped, subcloned, and partially sequenced. The insert (6.1 kb) contained sequence corresponding to a partially spliced mRNA species, and a 350-bp fragment was isolated from the exonic segment and used subsequently to screen an 8.5-d.p.c. mouse embryonic cDNA library. Two overlapping clones, Gli2-2 (2.1 kb)and Gli2-3 (1.9 kb), were isolated, and a composite sequence was established.

A single long ORF extends from nucleotides 110 to 4742 (the AUG at position 110 conforms to Kozak's rules for an initiating methionine; Kozak, 1984), predicting a protein of 1544 amino acids and 165 kDa. The protein contains five zinc finger motifs of the $C_2H_2$ subclass. The previously published *GLI2* sequence derives from human genomic fragments containing two exons corresponding to part of zinc finger 1 and zinc fingers 2–4 and a portion of domain 1 (Ruppert *et al.,* 1988). The amino acid sequence of the zinc fingers of

**TABLE 2**

**Alignment of Mouse GLI2 and Human GLI3 Protein Sequences**

| GLI2 | GLI3 | Identities/ total | % Identity | GLI domain[a] |
|---|---|---|---|---|
| 44–151 | 121–229 | 65/110 | 60 | |
| 212–284 | 279–348 | 54/70 | 77 | 1 |
| 293–393 | 361–454 | 51/101 | 54 | |
| 398–582 | 462–645 | 164/185 | 89 | 2 |
| 638–684 | 704–748 | 25/47 | 55 | 3 |
| 783–811 | 844–871 | 22/28 | 79 | 4 |
| 812–856 | 872–913 | 34/43 | 81 | |
| 856–885 | 914–943 | 25/30 | 83 | 5 |
| 886–925 | 944–982 | 23/41 | 60 | |
| 953–975 | 1003–1026 | 16/24 | 70 | |
| 1014–1020 | 1066–1972 | 6/7 | 85 | 6 |
| 1459–1476 | 1495–1512 | 13/18 | 72 | 7 |
| 1477–1512 | 1513–1548 | 26/36 | 72 | |

*Note.* The predicted mouse *Gli2* and human *GLI3* open reading frames were aligned using the GCG program bestfit.

[a] These domains are numbered according to the designation of Ruppert *et al.* (1990).

mouse *Gli2* has also been reported (Hui *et al.,* 1994). The mouse *Gli2* cDNA sequence displays 91% nucleotide and 100% amino acid identity with the human *GLI2* zinc finger nucleotide and amino acid sequences, respectively, 86% nucleotide and 100% amino acid identity with domain 1 (Ruppert *et al.,* 1988), and absolute identity with the published mouse *Gli2* zinc finger amino acid sequence (Hui *et al.,* 1994), confirming that we have identified the mouse *Gli2* gene. Searches of DNA and protein databases revealed marked homology with the other members of the *GLI* gene family as well as homology with the *Krüppel* and *TFIIIA* subclass of zinc finger containing proteins. GLI2 (mouse) and GLI3 (human) show the strongest homology (66% identity at the nucleic acid and 48% at the amino acid level). Previously Ruppert *et al.* (1988) defined seven domains of homology between human GLI and GLI3; these domains are conserved and extended when GLI2 and GLI3 are compared (Table 2). There are also additional regions of homology between GLI2 and GLI3 located N-terminal of region 1 and between domains 1 and 2 (Table 2).

An unexpected consequence of searching the databases with the mouse *Gli2* sequence was the finding of significant homology with human *THP/GLI4* (Tanimura *et al.,* 1993; Yoshida, 1995). *THP1* and *2* were originally isolated on the basis of binding to the TRE-2S sequence located in the long terminal repeat of HTLV-1, and the homology to the *Gli/Krüppel* family of genes was noted (Tanimura *et al.,* 1993). As a consequence of the absence of additional regions of shared homology with the *GLI* genes, Tanimura *et al.* (1993) suggested that *THP1* and *2* represent alternate splice forms of a new member of the *GLI* family, subsequently referring to it as *GLI4/THP* (Yoshida, 1995). However, *THP1* and *2* display absolute identity with the fragment of human *GLI2* zinc finger sequence available (data not shown), and comparison of the *THP1* and *2* nucleotide sequences with the mouse *Gli2* sequence reveals extensive homology outside the zinc finger region (Fig. 1a). The predicted translation product from nucleotides 431 to 583 (which corresponds to the translation initiation codon identified by Tanimura *et al.,* 1993) shows absolute identity with the corresponding mouse GLI2 amino acid sequence (Fig. 1b). There are also two single nucleotide gaps in the *THP1* and *2* sequence when lined up with the mouse *Gli2* sequence (equivalent to *Gli2* nucleotides 2320 and 2334, respectively); as a consequence of this an ORF with homology to that of *Gli2* changes frame twice, with the loss of the predicted stop codon for *THP1* and *2*. This ORF continues to the end of the available *THP* sequence. Therefore it is likely that *THP1* and *2* represent human *GLI2* and not a new member of the *GLI* family.

### *Gli2 Is Assigned to Mouse Chromosome 1*

Human *GLI2* has previously been assigned to chromosome 2 by hybridization to a human–rodent somatic cell hybrid panel (Ruppert *et al.,* 1988) that shows synteny with mouse chromosome 1 (DeBry and Seldin, 1996). We initially localized *Gli2* to band E of chromosome 1 by *in situ* hybridization to metaphase spreads of mouse chromosomes (data not shown).

To enable more precise localization, a microsatellite was isolated from a *Gli2* genomic clone. This locus was highly polymorphic across a number of inbred mouse strains including C57BL/6, DBA, and C3H (Fig. 2a); thus, we were able to map it on the BXD and BXH recombinant inbred strains, between *Emv17* and *En1* (Figs. 2b and 3b). We and others have previously demonstrated that the mouse mutation *Dh* maps to this region of chromosome 1 (Higgins *et al.,* 1992; Martin *et al.,* 1990). The *Dh* allele of *Gli2* was identified by analysis of the parents and offspring of a *Dh*/+ × C57BL/6 cross (Fig. 2a). Analysis of the relevant samples (8/561) from the *Dh* intraspecific backcross (Higgins *et al.,* 1992) demonstrated that this microsatellite and therefore *Gli2* was nonrecombinant with *Dh* (Fig. 3).

### *Candidacy of Gli2 for Dh*

*Gli2* represents a plausible candidate for the *Dh* mutation. The related gene *Gli3* has already been shown to be mutated in both *Xt* and GCPS, in which the limbs are affected (Hui and Joyner, 1993; Vortkamp *et al.,* 1991). Both *Gli2* and *Gli3* are expressed in the developing embryo from 8.5–16.5 d.p.c. in a similar range of tissues (Hui and Joyner, 1993; Hui *et al.,* 1994). *Gli2* is expressed in those tissues affected in *Dh,* including the splanchnic component of lateral mesoderm (from which the spleen develops), the gut mesenchyme, and the prevertebrae and developing limb (Searle, 1964; Green, 1967; Hui and Joyner, 1993; Hui *et al.,* 1994). At least two of the *Xt* mutations described to date are believed to be null alleles, resulting in haploinsufficiency in the heterozygote (Hui and Joyner, 1993; Vortkamp *et al.,* 1992). Southern blot analysis of DNA from +/+ and *Dh*/+ adult mice failed


**FIG. 1.** Similarity between mouse *Gli2* and human *THP*. (**a**) Nucleotide sequence homology between mouse *Gli2* (this study) and human *THP* (Tanimura *et al.*, 1993). (**b**) Alignment of predicted translation products of mouse *Gli2* and human *THP*. The *THP* sequence was adjusted at nucleotides 2320 and 2334. The initiating methionine described by Tanimura *et al.* (1993) is higlighted in boldface.

to demonstrate any obvious deletions or rearrangements (data not shown). To determine whether *Gli2* was expressed in *Dh* mice, a *Dh*/+ × *Dh*/+ mating was set up, and embryos were removed at 10.5 days. DNA and RNA were isolated, and each embryo was genotyped on the basis of the microsatellite. RT-PCR with primer pair 511/512 was carried out on two samples for each of the three genotypes. The levels of amplification of *Gli2* were similar in all three genotypes for aliquots taken between 25 and 40 cycles (data not shown), suggesting that *Gli2* expression is unaffected in *Dh*. To confirm that *Gli2* is expressed in *Dh* homozygotes and heterozygotes, RT-PCR was performed with primer pair 12a/1b, encompassing a coding sequence variation that is detectable as a restriction site polymorphism (nucleotide 976; Table 3). The *Dh* allele of *Gli2* is digested with *Hpa*II, whereas the C57BL/6 allele

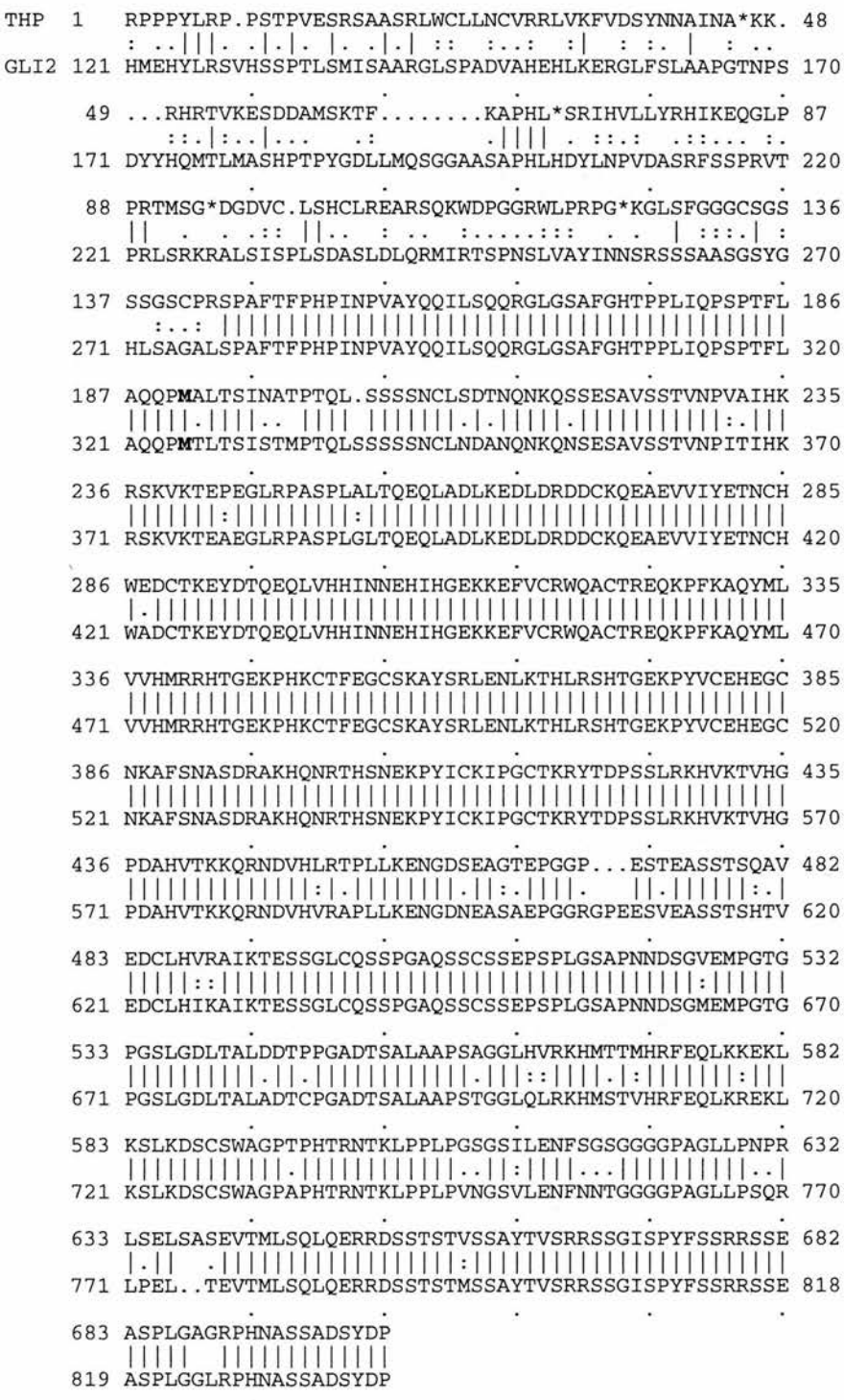


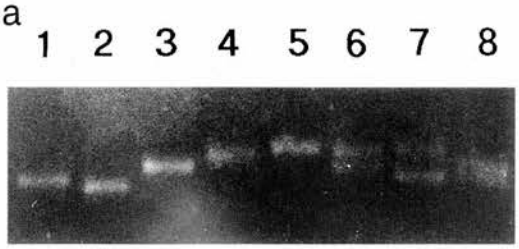

**FIG. 1**—*Continued*

is not. Thus we were able to demonstrate that *Gli2* is expressed in all three genotypes and both alleles are expressed in *Dh* heterozygotes (Fig. 4). The stronger signal observed for the C57BL/6-sized allele may represent the contribution of either undigested *Dh* homoduplex or, alternatively, a heteroduplex of both alleles.

*Analysis of Gli2 Coding Sequence*

In the absence of gross genomic abnormalities and changes in expression levels of *Gli2* in *Dh,* we analyzed the coding region for mutations. PCR primers were derived from the *Gli2* cDNA sequence to facilitate amplification of the entire coding sequence as a series of overlapping fragments (Table 1). RT-PCR generated fragments of the expected size for each region from each of the three genotypes (data not shown). The mismatch cleavage technique was used to look for any base changes (Condie *et al.,* 1993). An anomalous fragment was observed in fragment 1, and sequence analysis of PCR products identified a silent base change in both

a




b

| | 1 | 2 | 5 | 6 | 8 | 9 | 11 | 12 | 13 | 14 | 15 | 16 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 27 | 28 | 29 | 30 | 31 | 32 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *D1Mit11* | U | B | D | D | B | B | D | B | D | B | D | B | D | B | D | B | D | D | D | D | B | D | B | U | U | U |
| *D1Mit135* | U | B | D | D | B | B | D | B | D | B | D | B | D | B | D | B | D | D | D | B | B | D | B | D | B | D |
| *D1Mit163* | B | B | D | D | B | B | D | B | D | B | D | U | D | B | D | B | D | D | D | B | B | D | B | D | B | D |
| *D1Mit87* | B | B | D | D | B | B | D | B | D | B | D | D | D | B | D | B | D | D | D | B | B | D | B | D | B | D |
| *Emv17*[a] | B | B | D | D | B | B | D | B | D | B | D | D | D | B | D | B | D | D | D | B | B | D | D | D | B | D |
| *D1Mit417* | B | B | D | D | B | B | D | B | D | B | D | D | D | B | D | B | D | D | D | B | B | D | D | D | B | D |
| *D1Mit258* | B | B | U | D | B | B | D | B | D | B | D | D | D | B | D | B | D | D | D | B | B | D | D | D | B | D |
| *D1Mit285* | B | B | D | D | B | B | D | B | D | B | D | D | D | B | D | B | D | D | D | B | B | D | D | D | B | D |
| *Gli2* | B | B | D | D | B | B | D | B | D | B | D | D | D | B | D | B | B | D | D | B | B | D | D | D | B | D |
| *D1Ncvs44*[b] | B | B | D | D | B | B | D | B | D | B | D | D | D | B | D | B | B | D | D | B | B | D | D | D | B | D |
| *D1Ncvs45*[b] | B | B | D | D | B | B | D | B | D | U | D | D | D | B | D | B | B | D | D | B | U | U | D | D | B | D |
| *D1Mit188* | B | B | D | D | B | B | D | B | D | B | D | D | D | B | D | B | B | D | D | B | B | U | D | D | B | D |
| *D1Mit389* | B | B | D | D | B | B | D | B | D | B | D | D | D | B | D | B | B | D | D | B | D | D | D | D | B | D |
| *D1Mit419* | B | B | D | D | B | B | D | B | D | B | D | D | D | B | D | B | B | D | D | B | D | D | D | D | B | D |
| *D1Mit217* | B | B | D | D | B | B | D | B | D | B | D | D | D | B | D | B | B | D | D | B | D | D | D | D | B | D |
| *En1*[c] | B | B | D | D | B | B | D | B | D | B | D | D | D | B | D | B | B | D | D | B | D | D | D | D | B | D |
| *D1Mit88* | B | B | D | D | B | B | D | B | D | B | D | D | D | B | D | B | B | D | D | B | D | D | D | D | B | D |

| | 2 | 3 | 4 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 14 | 19 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Emv17*[a] | H | H | H | H | H | B | H | H | H | H | H | H |
| *Gli2* | B | H | H | H | H | B | U | U | H | H | H | H |
| *D1Mit217* | U | H | H | H | H | B | H | H | H | H | H | H |
| PEP3[a] | B | H | B | H | H | H | H | H | B | B | H | B |

**FIG. 2.** (**a**) PCR amplification of *Gli2*-associated microsatellite. DNA samples from a number of inbred mouse strains were amplified for the *Gli2*-associated microsatellite and electrophoresed on a 3% agarose gel. Lane **1,** C57BL/6; lane **2,** akr; lane **3,** 129; lane **4,** DBA; lane **5,** BALB/c; lane **6,** parental *Dh*/+; lane **7,** *Dh*/+ × C57BL/6 F1 (phenotype *Dh*/+); lane **8,** *Dh*/+ × C57BL/6 F1 (phenotype +/+). *Dh*/+ parents were heterozygous for the microsatellite. Analysis of F1 offspring identified the *Dh* allele as the larger. (**b**) Location of *Gli2* on mouse chromosome 1. The BXD and BXH RI strains were genotyped for the *Gli2*-associated microsatellite. Additional markers were placed on the map in this study, apart from [a](Buchberg *et al.,* 1986), [b](Hayashizaki *et al.,* 1994), and [c](Hill *et al.,* 1987).

DNA and RNA (nucleotide 976; Table 3). This sequence variation was also present in a number of inbred strains (data not shown). Additional variation was identified by sequencing of RT-PCR fragments from *Dh*/*Dh* and +/+ embryos, of which two sequence variations resulted in amino acid substitutions (Table 3). As *Dominant hemimelia* arose in a male mouse derived from CBA, C3H, or 101 (A. G. Searle, pers. comm., Harwell, 1994), we sequenced the relevant RT-PCR product from these inbred strains (data not shown). All three strains shared the same sequence as the *Dh* allele of *Gli2* at each variant. The complete coding sequence of *Gli2* was determined for *Dh,* and no mutations were identified.

*Genetic Map of the Dh Critical Region*

A number of microsatellite markers have been provisionally assigned to this region of mouse chromosome 1 (Dietrich *et al.,* 1992, 1994), so we sought to incorporate them into our genetic map around *Dh.* Markers were placed on both the BXD and BXH RI strains (Fig. 2b) and the *Dh* recombinant panel (Fig. 3a). The representation of these data is shown in Fig. 3b.

*Delineation of the Gli Family*

A number of genes have been described as members of the *GLI* family on the basis of homology in the zinc finger region, including the mammalian genes *YY1* (Shi *et al.,* 1991; Park and Atchison, 1991; Hariharan *et al.,* 1991; Flanagan *et al.,* 1991) and *Zic* (Aruga *et al.,* 1994) and the *Drosophila melanogaster odd paired* gene (Bendyk *et al.,* 1994). However, in all cases the homology does not extend beyond the zinc finger region. In addition, it has been proposed that there is a region of paralogy between mouse chromosomes 1 and 5 that includes the two homologs of the *Drosophila engrailed* gene, *En1* and *En2* (Martin *et al.,* 1990), and the homeobox genes *Gbx1* and *Gbx2* (Frohman *et al.,* 1993). It has been suggested that *Dh* and *Hammertoe,* and/or *Hemimelic Extra toe* may also be paralogous genes (Martin *et al.,* 1990). As *Gli* and *Gli3* have already been mapped to mouse chromosomes 10 and 13, respectively

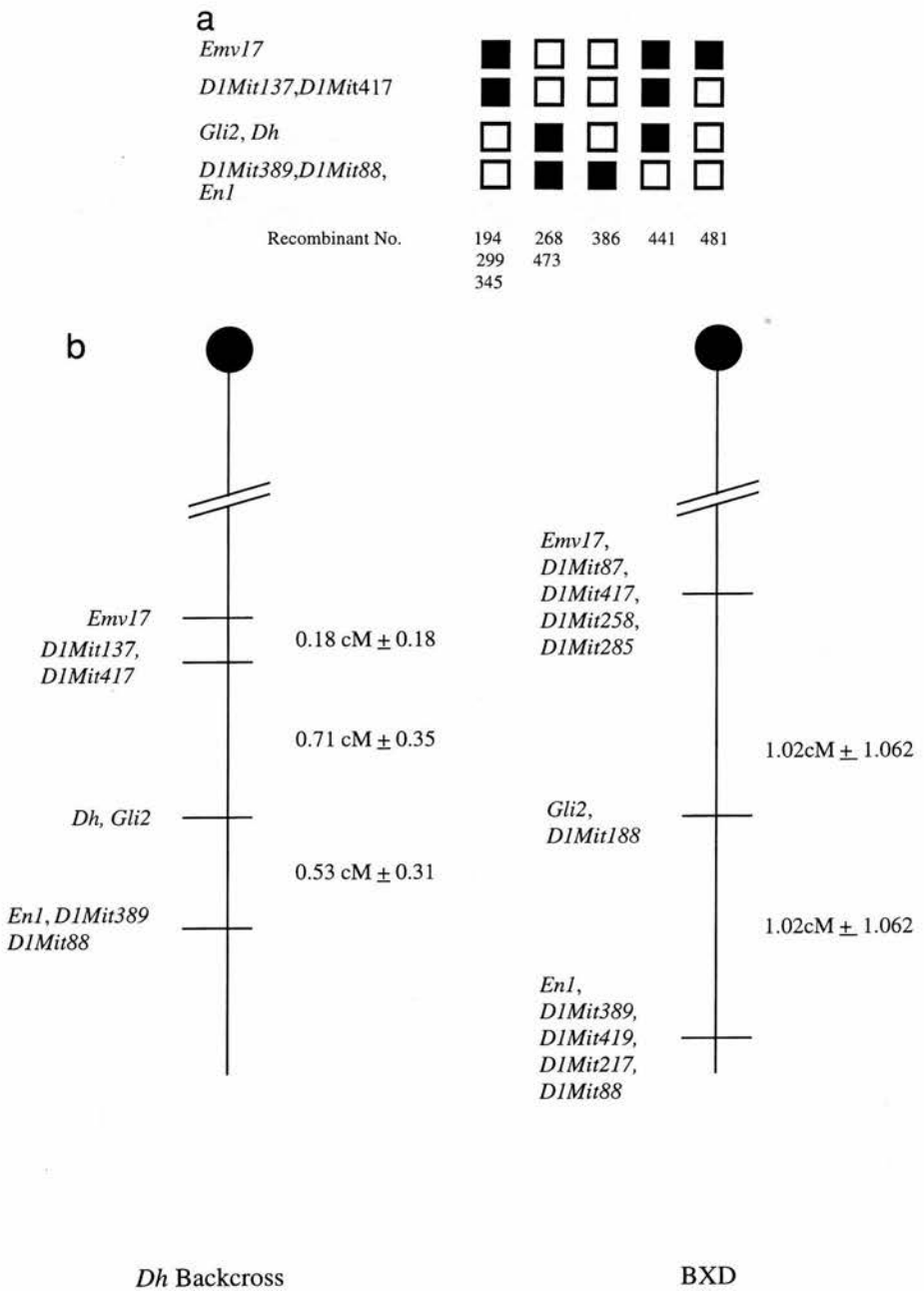



**FIG. 3.** (**a**) Cosegregation of *Dh* and *Gli2* on mouse chromosome 1 in a *Dh* intraspecific cross. Each column represents the chromosome that was inherited from the (*Dh*/+ × C57BL/6) F1 parent. Black boxes represent the *Dh* allele and white boxes the C57BL/6 allele. Data for additional markers were from Higgins *et al.* (1992) and this study. (**b**) Genetic linkage maps of mouse chromosome 1 in the vicinity of *Dh* derived from the *Dh* intraspecific backcross and the BXD RI strains. Loci are listed on the left side of the map. The right side of the map gives the estimated distances in centimorgans (cM) ± standard error.

(Justice *et al.,* 1990; Hui and Joyner, 1993), this region of paralogy may also include a novel *Gli*-like gene. We decided to test this proposition by determining whether there were any additional *Gli* genes in the mouse genome. We made use of a previously described "fingerprinting" strategy wherein amplification is carried out with redundant oligonucleotide primers based on conserved protein domains (Boehm, 1993).

Degenerate PCR primers were designed from the highly conserved zinc finger regions of *GLI, Gli2,* and *GLI3* (Kinzler *et al.,* 1988; Ruppert *et al.,* 1988; this study) to amplify between zinc fingers 2 and 5 (Table 1). Care was taken to avoid sequences also present in mammalian *Krüppel*-like zinc fingers, which may be found in at least 300 genes (Bellefroid *et al.,* 1989). RT-PCR was performed on mouse embryonic RNA (E12.5) with primer pair E187/E197, and only the expected fragment of 318 bp was observed (data not shown). Polyacrylamide gel electrophoresis of the digested PCR product revealed the presence of multiple DNA species (Table 4). Individual restriction fragments were isolated and ligated to a linker for subsequent amplification and cloning. Individual clones were sequenced and were found to correspond (in approximately equal pro-

**TABLE 3**

**Sequence Variation of the Mouse *Gli2* Gene**

| Nucleotide position | C57BL/6 | *Dh* | Amino acid position | C57BL/6 | *Dh* |
|---|---|---|---|---|---|
| 301 | C | A | 44 | His | Gln |
| 446 | G | T | 113 | Ala | Thr |
| 976 | T | G | NA | | |
| 992 | A | C | NA | | |
| 5410 | $(G)_{12}$ | $(G)_8$ | NA | | |

*Note.* Nucleotide sequence variation between C57BL/6 and *Dh* alleles of the mouse *Gli2* gene and corresponding amino acid changes (where applicable) are shown. Each sequence variation was also found in C3H, 101, and CBA inbred mouse strains. The numbering is based on the *Gli2* cDNA sequence.

portions) only to *Gli, Gli2,* and *Gli3* (Table 4), suggesting that there are no additional expressed *Gli* genes. However, such a gene may not be expressed at the embryonic stage tested or may display slight sequence divergence at the primer binding site(s), preventing successful amplification. We therefore designed degenerate primers to amplify a different region of the zinc fingers from genomic DNA, i.e., from within a single exon (E195/E196; Table 1). The expected fragment of 123 bp was amplified from mouse DNA and subcloned. Individual clones were sequenced and found to correspond to *Gli* ($n = 10$) and *Gli2* ($n = 6$). No clones corresponding to *Gli3* or any novel *Gli* genes were identified. The amplification was repeated with an initial series of cycles at lower stringency. Thirty-two clones were sequenced representing all three *Gli* genes (Table 4) and one artifact (data not shown). Therefore, using two sets of primers to different regions of the highly conserved *Gli* zinc finger domain, we were able to identify only the previously described genes. It is unlikely that there are any additional *Gli* genes in the mouse genome.

## DISCUSSION

The *GLI* family represents a functionally distinct, evolutionarily conserved group of zinc finger genes (Ruppert *et al.,* 1988; Orenic *et al.,* 1990; Zarkower and Hodgkin, 1992; Holland *et al.,* 1991) despite similarity to the *Krüppel* family of genes. Although the human *GLI* gene was identified because it was amplified in a human glioma (Kinzler *et al.,* 1987), its role in the neoplastic process is unclear (Reifenberger *et al.,* 1994). All three mammalian *GLI* genes are expressed in various ectoderm- and mesoderm-derived tissues from 8.5–16.5 d.p.c. in the mouse, suggesting that they play multiple roles during postimplantation development. The role of *Gli3* in the development of a number of tissues has been established by the finding that haploinsufficiency for *GLI3* is believed to be the causative factor in both GCPS and the mouse *Xt* mutation. All the semidominant *Xt* alleles exhibit homozygous lethality and display a greater range of abnormalities compared to heterozygotes (Hui and Joyner, 1993). While two of the *Xt* alleles are due to partial deletion of *Gli3,* in a third allele $Xt^{Pdn}$, *Gli3* is not deleted nor is there any diminution in *Gli3* expression (Schimmang *et al.,* 1994); $xt^{add}$ is a recessive allele where the integration of a transgene upstream of *Gli3* results in a decrease in the expression of *Gli3* (Van der Hoeven *et al.,* 1993). There is some evidence that an ectopic zone of polarizing activity is generated in the limb bud of *Xt* mice and that this may contribute to the limb phenotype (Masuya *et al.,* 1995).

The conservation of peptide domains within the mammalian GLI proteins and indeed compared to the *Drosophila* CI protein suggests that these are either structural domains or sites of interaction with other proteins. The GLI proteins have been proposed as transcription factors on the basis of nuclear localization (Kinzler and Vogelstein, 1990) and DNA binding activity (Kinzler and Vogelstein, 1990; Ruppert *et al.,* 1990; Vortkamp *et al.,* 1995). The level of homology within the zinc finger region, particularly fingers 3–5, predicts that the GLI proteins interact with the same or similar target DNA sequences, and it has been demonstrated that GLI, GLI3, and TRA-1 bind very similar sequences *in vitro* (Kinzler and Vogelstein, 1990; Ruppert *et al.,* 1990; Vortkamp *et al.,* 1995; Zarkower and Hodgkin, 1993). Pavletich and Pabo (1993) have demonstrated that fingers 4 and 5 of GLI have the main base contacts to the 9-bp consensus sequence. However, despite the identification of consensus DNA target sequences, the *in vivo* targets of the GLI proteins have yet to be identified (Kinzler and Vogelstein, 1990; Vortkamp *et al.,* 1995).

The observation that the TRE-2S holding protein (THP) is very probably human GLI2 may provide some clues to how the GLI proteins function. The Tax protein of HTLV-1 efficiently transactivates two or more copies of the 21-bp enhancer (TRE-1) but fails to transactivate one copy (Tanimura *et al.,* 1993). However, an adjacent sequence (TRE-2S) is sufficient to enhance the activity of one copy of the enhancer, and THP was identified because it binds to the TRE-2S sequence (Tanimura *et al.,* 1993). Tanimura *et al.* (1993) demonstrated that

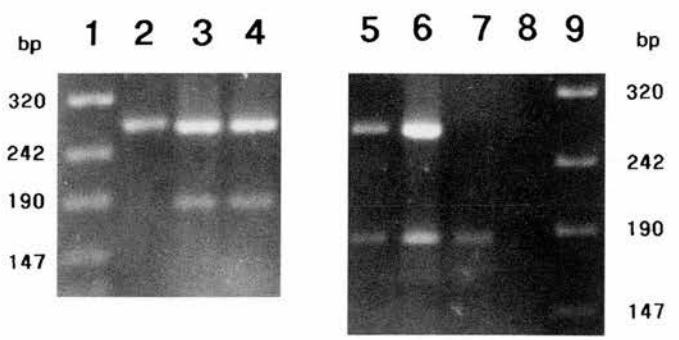



**FIG. 4.** Expression of *Gli2* in *Dh.* Embryonic RNA was obtained from all three genotypes, and a *Gli2*-specific fragment (12a/1b) was generated by RT-PCR. Following digestion with *Hpa*II, products were electrophoresed on 3% agarose gels. Lanes **1** and **9,** molecular weight marker VIII (Boehringer Mannheim); lane **2,** +/+; lanes **3–6,** *Dh*/+; lane **7,** *Dh*/*Dh;* lane **8,** −RNA control.

**TABLE 4**

**Determination of the Number of *Gli* Genes in the Mouse Genome**

| Source of template | | | | | | |
|---|---|---|---|---|---|---|
| cDNA | | | | Genomic DNA | | |
| Restriction enzyme | Fragment size (bp) | Number of clones sequenced | Identity of sequenced clones | Identity of clones | Number of clones sequenced (normal stringency PCR) | Number of clones sequenced (low stringency PCR) |
| *Alu*I | 280 | 0 | NA | *Gli* | 10 | 10 |
| | 249 | 2 | *Gli* (2) | *Gli2* | 6 | 10 |
| *Hae*III | 192 | 2 | *Gli3* (2) | *Gli3* | 0 | 11 |
| | 180 | 2 | *Gli* (2) | | | |
| | 83 | 3 | *Gli2* (3) | | | |
| *Msp*I | 318 | 1 | *Gli2* (1) | | | |
| | 288 | 2 | *Gli3* (2) | | | |
| | 254 | 2 | *Gli* (2) | | | |
| *Rsa*I | 318 | 1 | *Gli3* (1) | | | |
| | 269 | 4 | *Gli* (3) | | | |
| | | | *Gli2* (1) | | | |
| *Sau*3AI | 275 | 3 | *Gli2* (3) | | | |
| | 269 | 3 | *Gli* (1) | | | |
| | | | *Gli2* (2) | | | |
| | 194 | 2 | *Gli3* (2) | | | |
| | 120 | 2 | *Gli3* (2) | | | |
| *Taq*I | 90 | 3 | *Gli3* (3) | | | |
| | 67 | 0 | NA | | | |

TRE-2S has no intrinsic enhancer activity, but acts in concert with a single copy of the TRE-1 enhancer; yet they could not demonstrate this cooperative transactivation when the *THP* clone was cotransfected with a Tax expression plasmid. As we have demonstrated, though, the reported *THP* sequence encodes a severely truncated GLI2 protein that would be capable of DNA binding, but lacking domains 1, 5, 6, 7, and 8, would probably be deficient in a number of protein–protein interactions.

There is also evidence of a requirement for additional factors in the functioning of GLI3. Vortkamp *et al.* (1995) failed to show significant activation or repression of basal expression levels using a full-length *Gli3* clone and triplicate *Gli3* binding sites in transient transfection assays in NIH 3T3 cells. This is not surprising if the GLI proteins are not sufficient to regulate transcription, but in fact enhance transcription via other factors. If the GLI proteins do enhance the activity of other transcription factors, then it is possible that they interact with CREB and/or CREM (Tanimura *et al.*, 1993). Mutations in the human CREB binding protein (*CBP*) gene have been described in Rubinstein–Taybi syndrome, which is characterized by facial abnormalities and broad thumbs (Petrij *et al.*, 1995), and long-range patterning by sonic hedgehog is blocked by increases in cAMP (Fan *et al.*, 1995). Thus the GLI proteins may cooperate with cAMP signaling pathways in the development of the mammalian limb.

We have localized *Gli2* to the same region of mouse chromosome 1 as *Dominant hemimelia*. The first defect observed in *Dh* is disruption of the splanchnic mesoderm (Green, 1967); the spleen forms in the region where the anterior splanchnic mesodermal plate develops. Absence of the spleen is not itself sufficient to explain the range of effects observed in *Dh*. The spleen fails to develop in mice that are homozygous for null alleles of *Hox11*, yet there is no evidence of any additional abnormalities (Roberts *et al.*, 1994, 1995; Dear *et al.*, 1995); it is not clear in these mutant mice whether the spleen fails to develop or whether spleen development is initiated but then fails due to apoptosis at 13.5 d.p.c. (Roberts *et al.*, 1995; Dear *et al.*, 1995). *Gli2* is expressed in those tissues that are affected in *Dh* such as the splanchnic component of lateral mesoderm, developing limb bud, and gut mesenchyme (Hui *et al.*, 1994), and therefore *Gli2* represents a very good candidate for the *Dh* mutation. However, we have not demonstrated any *Gli2* coding sequence mutations, suggesting that *Gli2* is not mutated in *Dh*. We cannot rule out mutations in the regulatory regions of the *Gli2* gene that may have subtle effects on the expression of *Gli2*.

## ACKNOWLEDGMENTS

We thank Neil Brockdorf for help with the lung cDNA library screening and Kathryn Newton for technical assistance. We thank Karen Steel for her comments on the manuscript. We are grateful to Brigid Hogan for the gift of the mouse embryo cDNA library. K.S. is a Royal Society "Japan Return Fellow."

## REFERENCES

Aruga, J., Yokota, N., Hashimoto, M., Furuichi, T., Fukuda, M., and Mikoshiba, K. (1994). A novel zinc-finger protein, *zic*, is involved

in neurogenesis, especially in the cell lineage of cerebellar granule cells. *J. Neurochem.* **63:** 1880–1818.

Bellefroid, E. J., Lecocq, P. J., Benhida, A., Poncelet, D. A., Belayew, A., and Martial, J. A. (1989). The human genome contains hundreds of genes coding for finger proteins of the *Krüppel* type. *DNA* **8:** 377–387.

Benedyk, M. J., Mullen, J. R., and Dinardo, S. (1994). *Odd-paired*—A zinc-finger pair rule protein required for the timely activation of *engrailed* and *wingless* in *Drosophila* embryos. *Genes Dev.* **1:** 105–117.

Boehm, T. (1993). Analysis of multigene families by DNA fingerprinting of conserved domains—Directed cloning of tissue specific protein tyrosine phosphatases. *Oncogene* **8:** 1385–1390.

Buchberg, A. M., Taylor, B. A., Jenkins, N. A., and Copeland, N. G. (1986). Chromosomal localization of *Emv-16* and *Emv-17,* two closely linked ecotrpic proviruses of RF/J mice. *J. Virol.* **60:** 1175–1178.

Condie, A., Eeles, R., Borresen, A. L., Coles, C., Cooper, C., and Prosser, J. (1993). Detection of point mutations in the P53 gene—Comparison of single-strand conformation polymorphism, constant denaturant gel-electrophoresis, and hydroxylamine and osmium tetroxide techniques. *Hum. Mut.* **2:** 58–66.

Copeland, N. G., Gilbert, D. J., Jenkins, N. A., Nadeau, J. H., Eppig, J. T., Maltais, L. J., Miller, J. C., Dietrich, W. F., Steen, R. G., Lincoln, S. E., Weaver, A., Joyce, D. C., Merchant, M., Wessel, M., Katz, H., Stein, L. D., Reeve, M. P., Daly, M. J., Dredge, R. D., Marquis, A., Goodman, N., and Lander, E. S. (1993). Genome Maps IV. *Science* **262:** 67.

Dear, T. N., Colledge, W. H., Carlton, M. B. L., Lavenir, I., Larson, T., Smith, A. J. H., Warren, A. J., Evans, M. J., Sofroniew, M. V., and Rabbitts, T. H. (1995). The *Hox11* gene is essential for cell survival during spleen development. *Development* **121:** 2909–2915.

Debry, R. W., and Seldin, M. F. (1996). Human/mouse homology relationships. *Genomics* **33:** 337–351.

Dietrich, W., Katz, H., Lincoln, S. E., Shin, H. S., Friedman, J., Dracopoli, N. C., and Lander, E. S. (1992). A genetic map of the mouse suitable for typing intraspecific crosses. *Genetics* **131:** 423–447.

Dietrich, W. F., Miller, J. C., Steen, R. G., Merchant, M., Damron, D., Nahf, R., Gross, A., Joyce, D. C., Wessel, M., Dredge, R. D., Marquis, A., Stein, L. D., Goodman, N., Page, D. C., and Lander, E. S. (1994). A genetic map of the mouse with 4,006 simple sequence length polymorphisms. *Nature Genet.* **7:** 220–245.

Fan, C. M., Porter, J. A., Chiang, C., Chang, D. T., and Beachy, P. A. (1995). Long range sclerotome induction by sonic hedgehog-direct role of the amino-terminal cleavage product and modulation by the cyclic AMP signalling pathway. *Cell* **81:** 457–465.

Flanagan, J. R., Becker, K. G., Ennist, D. L., Gleason, S. L., Driggers, P. H., Levi, B.-Z., Apella, E., and Ozato, K. (1991). Cloning of a negative transcription factor that binds to the upstream conserved region of Moloney murine leukaemia virus. *Mol. Cell. Biol.* **12:** 38–44.

Frohman, M. A., Dickinson, M. E., Hogan, B. L. M., and Martin, G. R. (1993). Mapping of *GBX-1* to mouse chromosome 5 and *GBX-2* to mouse chromosome 1. *Mouse Genome* **91:** 323–325.

Green, M. C. (1967). A defect of the splanchnic mesoderm caused by the mutant gene *Dominant hemimelia* in the mouse. *Dev. Biol.* **15:** 62–89.

Hariharan, N., Kelley, D. E., and Perry, R. P. (1991). Delta, a transcription factor that binds to downstream elements in several polymerase II promoters, is a functionally versatile zinc finger. *Proc. Natl. Acad. Sci. USA* **88:** 9799–9803.

Higgins, M., Hill, R. E., and West, J. D. (1992). *Dominant hemimelia* and *En-1* on mouse chromosome 1 are not allelic. *Genet. Res. Cambridge* **60:** 53–60.

Hill, R. E., Hall, A. E., Sime, C. M., and Hastie, N. D. (1987). A mouse homeobox-containing gene maps near a developmental mutation. *Cytogenet. Cell Genet.* **44:** 171–174.

Holland, P. W. H., Williams, N. A., and Lanfear, J. (1991). Cloning of segment polarity gene homologues from the unsegmented brachiopod *Terebratulina retusa* (Linnaeus) *FEBS Lett.* **291:** 211–213.

Hui, C-C., and Joyner, A. L. (1993). A mouse model of Greig cephalopolysyndactyly syndrome: The *extra-toes*$^{J}$ mutation contains an intragenic deletion of the Gli3 gene. *Nature Genet.* **3:** 241–246.

Hui, C-C., Slusarski, D., Platt, K. A., Holmgren, R., and Joyner, A. L. (1994). Expression of three mouse homologs of the *Drosophila* segment polarity gene *cubitis interruptus, Gli, Gli-2,* and *Gli-3,* in ectoderm- and mesoderm-derived tissues suggest multiple roles during postimplantation development. *Dev. Biol.* **162:** 402–413.

Justice, M. J., Siracusa, L. D., Gilbert, D. J., Heisterkamp, N., Groffen, J., Chada, K., Silan, C. M., Copeland, N. G., and Jenkins, N. A. (1990). A genetic linkage map of mouse chromosome 10: Localization of eighteen molecular markers using a single interspecific backcross. *Genetics* **125:** 855–866.

Kawasaki, E. S. (1990). Amplification of RNA. *In* "PCR Protocols: A Guide To Methods And Applications" (M. A. Innis, D. H. Gelfand, J. J. Sninsky, and T. J. White, Eds.), Academic Press, San Diego.

Kinzler, K. W., Bigner, S. H., Bigner, D. D., Trent, J. M., Law, M. L., O'Brien, S. J., Wong, A. J., and Vogelstein, B. (1987). Identification of an amplified, highly expressed gene in human glioma. *Science* **236:** 70–73.

Kinzler, K. W., Ruppert, J. M., Bigner, S. H., and Vogelstein, B. (1988). The *GLI* gene is a member of the *Krüppel* family of zinc finger proteins. *Nature (London)* **332:** 371–374.

Kinzler, K. W., and Vogelstein, B. (1990). The GLI gene encodes a nuclear protein which binds specific sequences in the human genome. *Mol. Cell. Biol.* **10:** 634–642.

Kozak, M. (1984). Compilation and analysis of sequences upstream from the translational start site in eukaryotic mRNAs. *Nucleic Acids Res.* **12:** 857–871.

Martin, G. R., Richman, M., Reinsch, S., Nadeau, J. H., and Joyner, A. (1990). Mapping of the two mouse *engrailed*-like genes: Close linkage of *En-1* to *dominant hemimelia* (*Dh*) on Chromosome 1 and of *En-2* to *hemimelic extra-toes* (*Hx*) on Chromosome 5. *Genomics* **6:** 302–308.

Masuya, H., Sagai, T., Wakana, S., Moriwaka, K., and Shiroishi, T. (1995). A duplicated zone of polarizing activity in polydactylous mouse mutants. *Genes Dev.* **9:** 1645–1653.

Mattei, M. G., Philip, N., Passage, E., Moisan, J. P., Mandel, J. L., and Mattei, J. F. (1985). DNA probe localization at 18p11.3 band by in situ hybridization and identification of a supernumerary chromosome. *Hum. Genet.* **69,** 268–271.

Orenic, T., Slusarski, D. C., Kroll, K. L., and Holmgren, R. (1990). Cloning and characterisation of the segment polarity gene *cubitis interruptus Dominant* of *Drosophila. Genes Dev.* **4:** 1053–1067.

Park, K., and Atchison, M. L. (1991). Isolation of a candidate repressor/activator NF-E1 (YY1, delta), that binds to the immunoglobulin kappa 3-prime enhancer and the immunoglobulin heavy-chain micro-E1 site. *Proc. Natl. Acad. Sci. USA* **88:** 9804–9808.

Pavletich, N. P., and Pabo, C. O. (1993). Crystal structure of a 5-finger GLI-DNA complex—New perspectives on zinc fingers. *Science* **261:** 1701–1707.

Petrij, F., Giles, R. H., Dauwerse, H. G., Saris, J. J., Hennekam, R. C. M., Masuno, M., Tommerup, N., van Ommen, G. B., Goodman, R. H., Peters, D. J. M., and Breuning, M. H. (1995). Rubinstein–Taybi syndrome caused by mutations in the transcriptional co-activator CBP. *Nature* **376:** 348–351.

Reifenberger, G., Reifenberger, J., Ichimura, K., Meltzer, P. S., and Collins, V. P. (1994). Amplification of multiple genes from chromosomal region 12q13–14 in human malignant gliomas: Preliminary mapping of the amplicons shows preferential involvement of *CDK4, SAS,* and *MDM2. Cancer Res.* **54:** 4299–4303.

Roberts, C. W. M., Shutter, J. R., and Korsmeyer, S. J. (1994). *Hox11* controls the genesis of the spleen. *Nature* **368:** 747–749.

Roberts, C. W. M., Sonder, A. M., Lumsden, A., and Korsmeyer, S. J.

(1995). Developmental expression of *Hox11* and specification of splenic cell fate. *Am. J. Pathol.* **146:** 1089–1101.

Ruppert, J. M., Kinzler, K. W., Wong, A. J., Bigner, S. H., Kao, F-T., Law, M. L., Seuanez, H. N., O'Brien, S. J., and Vogelstein, B. (1988). The *GLI-Krüppel* family of human genes *Mol. Cell. Biol.* **8:** 3104–3113.

Ruppert, J. M., Vogelstein, B., Arheden, K., and Kinzler, K. W. (1990). GLI3 encodes a 190-Kilodalton protein with multiple regions of GLI similarity. *Mol. Cell. Biol.* **10:** 5408–5415.

Schimmang, T., Lemaistre, M., Vortkamp, A., and Ruther, U. (1992). Expression of the zinc finger gene *Gli3* is affected in the morphogenetic mouse mutant *extra-toes* (*Xt*). *Development* **116:** 799–804.

Schimmang, T., Oda, S.-I., and Ruther, U. (1994). The mouse mutant *Polydactyly Nagoya* (*Pdn*) defines a novel allele of the zinc finger gene *Gli3*. *Mamm. Genome* **5:** 384–386.

Searle, A. G. (1964). The genetics and morphology of two 'luxoid' mutants in the house mouse. *Genet. Res.* **5:** 171–197.

Shi, Y., Seto, E., Chang, L.-S., and Shenk, T. (1991). Transcriptional repression by YY1, a human GLI-Krüppel-related protein, and relief of repression by adenovirus E1A protein. *Cell* **67:** 377–388.

Tanimura, A., Teshima, H., Fujisawa, J-I., and Yoshida, M. (1993). A new regulatory element that augments the Tax-dependent enhancer of Human T-cell Leukaemia Virus Type 1 and cloning of cDNAs encoding its binding proteins. *J. Virol.* **67:** 5375–5382.

Van der Hoeven, F., Schimmang, T., Vortkamp, A., and Ruther, U. (1993). Molecular linkage of the morphogenetic mutation *add* and the zinc finger gene *Gli3*. *Mamm. Genome* **4:** 276–277.

Vortkamp, A., Gessler, M., and Grzeschik, K.-H. (1991). *GLI3* zinc-finger gene interrupted by translocations in Grieg syndrome families. *Nature* **352:** 539–540.

Vortkamp, A., Franz, T., Gessler, M., and Grzeschik, K.-H. (1992). Deletion of *GLI3* supports the homology of the human Grieg cephalopolysyndactyly syndrome (GCPS) and the mouse mutant *extra-toes* (*Xt*). *Mamm. Genome* **3:** 461–463.

Vortkamp, A., Gessler, M., and Grzeschik, K.-H. (1995). Identification of optimized target sequences for the GLI3 zinc finger protein. *DNA Cell Biol.* **14:** 629–634.

Winship, P. R. (1989). An improved method for directly sequencing PCR amplified material using dimethyl sulphoxide. *Nucleic Acids Res.* **17:** 1266.

Yoshida, M. (1995). HTLV-1 oncoprotein Tax deregulates transcription of cellular genes through multiple mechanisms. *J. Cancer Res. Clin. Oncol.* **121:** 521–528.

Zarkower, D., and Hodgkin, J. (1992). Molecular analysis of the *C.elegans* sex determining gene *tra-1:* A gene encoding two zinc finger proteins. *Cell* **70:** 237–249.

Zarkower, D., and Hodgkin, J. (1993). Zinc fingers in sex determination—Only one of the two *C.elegans tra-1* proteins binds DNA in vitro. *Nucleic Acids Res.* **21:** 3691–3698.